格致方法
定量研究系列

· 革新研究理念
· 丰富研究工具
· 最权威、最前沿的定量研究方法指南

吴晓刚 / 主编
【美】诺瓦尔 · D. 格伦（Norval D. Glenn）/ 等著

纵贯数据分析

格致出版社 上海人民出版社

世纪出版

格致方法
定量研究系列

纵贯数据分析

本书由四种讨论纵贯研究的小册子组成，分别是《世代分析》、《纵贯研究》、《分析重复调查数据》以及《多元时间序列模型》。纵贯研究相较截面研究而言，其最大的优点是可以描述事物的发展过程和变化，并从这种变化中考察社会发展趋势。本书分别对纵贯研究中极具代表性的几种分析进行了介绍，旨在使研究者对纵贯研究有一定的认识，并能以此方法进行研究。

上架建议：社科研究方法

ISBN 978-7-5432-1903-8

9 787543 219038 >

定价：55.00元

易文网：www.ewen.cc

格致网：www.hibooks.cn

格致方法
定量研究系列

吴晓刚 / 主编

【美】诺瓦尔·D. 格伦(Norval D. Glenn) / 等著

纵贯数据分析

SAGE Publications ,Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

纵贯数据分析/(美)格伦(Glenn，N. D.)等著；吴晓刚主编. —上海：格致出版社：上海人民出版社，2011

(格致方法·定量研究系列)

ISBN 978-7-5432-1903-8

Ⅰ. ①纵… Ⅱ. ①格… ②吴… Ⅲ. ①统计数据-统计分析(数学) Ⅳ. ①O212.1

中国版本图书馆CIP数据核字(2010)第254753号

责任编辑 顾 悦

封面装帧 人马艺术工作室·储平

纵贯数据分析

[美]诺瓦尔·D.格伦 等著

吴晓刚 主编

出 版 世纪出版集团 www.ewen.cc 格致出版社 www.hibooks.cn 上海人民出版社

(200001 上海福建中路193号24层)

编辑部热线 021-63914988

市场部热线 021-63914081

发 行 世纪出版集团发行中心

印 刷 上海图宇印刷有限公司印刷

开 本 787×1092毫米 1/16

印 张 21.5

插 页 1

字 数 309,000

版 次 2011年8月第1版

印 次 2011年8月第1次印刷

ISBN 978-7-5432-1903-8/C·48

定 价 55.00元



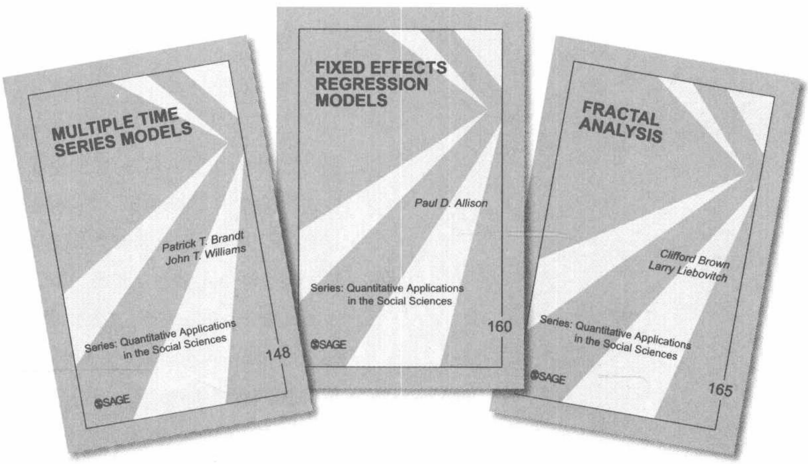
吴晓刚

美国加州大学洛杉矶分校(UCLA)社会学博士，香港科技大学社会科学部教授、香港科技大学应用社会经济研究中心主任。主要研究领域为社会分层和流动、劳动力市场和经济社会学、教育、社会人口学、统计学和定量研究方法。多篇论文发表于 *American Journal of Sociology*、*American Sociological Review*、*Social Forces* 等国际著名专业杂志上。2006 年获美国国家教育学院 Spencer 博士后研究奖学金；2007 年 “以精辟数据研究中国社会” 获美国社会学亚洲人 / 亚裔美国人 “青年学人” 奖。目前担任多家英文学术期刊的编委和 SSCI 期刊 *Chinese Sociological Review* (CSR) 的主编。

香港科技大学应用社会经济研究中心 (CASER) 成立于 2009 年 12 月。中心倡导以定量社会科学研究为主，兼容多种研究方法，结合多学科视角，对中国面临的重大社会经济问题及其对人民生活福祉的影响进行实证研究。中心的研究项目涉及教育、就业、不平等、贫困、移民、健康等等。中心主持了“香港社会动态追踪调查”，并与中国人民大学合作，参与“中国教育追踪调查”。中心强调研究的科学性、独立性和社会意义，并期望通过讨论会和工作坊以激发讨论，将研究成果与更多的公众分享。中心网站 : http://caser.ust.hk。

SAGE

绿皮书的传奇



1976 年，当政府管理与政治学教授 Eric Uslaner 向 SAGE 出版社的创始人 Sara 建议出版关于重要定量研究方法的简明小册子时，没有人预见到这套书会取得巨大成功。

那年夏天，"社会科学定量研究方法" 丛书诞生。统一的、朴素的绿色封面，上面仅仅印着书名、作者名及本书所属的系列名。Iverson 和 Northporth 的《变量分析》(*Analysis of Variance*)、Nagel 的《运筹学》(*Operations Research*) 以及 Henkel 的《显著性检验》(*Tests of Significance*) 是最早出版的几本，每本售价 2.95 英镑，被形容为 "像烤饼一样好卖"。

SAGE 选择了 20 种顶尖的定量研究工具，然后去寻找合适的作者，邀请他们围绕这些工具撰写 92 页的小书。这些薄薄的绿皮书在课堂上深受欢迎，在图书馆成套陈列。

至今，绿皮书系列中共有 160 种在初版或再版。书的主题反映了量化研究方法的发展：从基础统计知识、数据类型、测量到计算机的应用以及博弈论。这套书非常畅销，其中最畅销的一本，是 1980 年出版的 Michael Lewis-Beck 的《回归方法的应用》(*Applied Regression*)。

出版说明

本书由四种讨论纵贯研究的小册子组成，分别是《世代分析》、《纵贯研究》、《分析重复调查数据》以及《多元时间序列模型》。纵贯研究相较截面研究而言，其最大的优点是可以描述事物的发展过程和变化，并从这种变化中考察社会发展趋势。但是，纵贯研究的执行往往必须建立在多年的数据收集和分析的基础之上，而且研究的重复性较大。鉴于该研究具有较高的难度和复杂性，本书分别对纵贯研究中极具代表性的几种分析进行了介绍，旨在使研究者对纵贯研究有一定的认识，并能以此方法进行研究。

《世代分析》介绍了检验数据时常用的一项分析方法，即世代研究。这本小册子分别讨论了世代研究的目的、定义、策略等问题，并结合实例，说明了该研究的具体应用和数据选择问题。《纵贯研究》讨论了纵贯研究的目的、数据收集、研究对象与研究模型等问题，能为研究发展趋势、历史变迁或动态过程的研究者提供帮助。《分析重复调查数据》指出，重复调查是从对个人层次的微观研究转向对整体层次的宏观研究，并讨论了重复调查的基本用途，即描述、分解、解释总体趋势、对个别参数变化进行估计。《多元时间序列模型》概述了多元时间序列分析中的许多高级方法，包括同时方程模型、向量自回归模型、格兰杰因果检验、冲击反应分析等。作者通过介绍这些方法，预测并描述了变量之间内生的动态关系。

总序

往事如烟，光阴如梭。转眼间，出国已然十年有余。1996 年赴美留学，最初选择的主攻方向是比较历史社会学，研究的兴趣是中国的制度变迁问题。以我以前在国内所受的学术训练，基本是看不上定量研究的。一方面，我们倾向于研究大问题，不喜欢纠缠于细枝末节。国内一位老师的话给我的印象很深，大致是说：如果你看到一堵墙就要倒了，还用得着纠缠于那堵墙的倾斜角度究竟是几度吗？所以，很多研究都是大而化之，只要说得通即可。另一方面，国内（十年前）的统计教学，总的来说与社会研究中的实际问题是相脱节的。结果是，很多原先对定量研究感兴趣的学生在学完统计之后，依旧无从下手，逐渐失去了对定量研究的兴趣。

我所就读的美国加州大学洛杉矶分校社会学系，在定量研究方面有着系统的博士训练课程。不论研究兴趣是定量还是定性的，所有的研究生第一年的头两个学期必须修两门中级统计课，最后一个学期的系列课程则是简单介绍线性回归以外的其他统计方法，是选修课。希望进一步学习定量研究方法的可以在第二年修读另外一个三学期的系列课程，其中头两门课叫“调查数据分析”，第三门叫“研究设计”。除此以外，还有如“定类数据分析”、“人口学方法与技术”、“事件史分析”、“多层线性模型”等专门课程供学生选修。该学校的统计系、心理系、教育系、经济系也有一批蜚声国际的学者，提供不同的、更加专业化的课程供学生选修。2001 年完成博士学业之后，我又受安德鲁·梅隆基金会资助，在世界定量社会科学研究的重镇密歇根大学从事两年的博士后研究，其间旁听谢宇教授为博士生讲授的统计课程，并参与该校社会研究院（Institute for Social Research）定量社会研究方法项目的一些讨论会，受益良多。

2003年，我赴港工作，在香港科技大学社会科学部，教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来，非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课，而第二门课“社会科学中的定量分析”为博士生的必修课(事实上，大部分硕士生在修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候，根据社会科学研究生的数理基础比较薄弱的特点，尽量避免复杂的数学公式推导，而用具体的例子，结合语言和图形，帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上，即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”，学完这些课程后，我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章；另一方面，也能在自己的研究中运用这些成熟的方法论技术。

上述两门课的内容，尽管在线性回归模型的内容上有少量重复，但各有侧重。“社会科学里的统计学”(Statistics for Social Science)从介绍最基本的社会研究方法论和统计学原理开始，到多元线性回归模型结束，内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型，以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法，将重点放在因变量为定类数据的分析模型上，包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型，包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生，一直鼓励和支持我将两门课的讲稿结集出版，并帮助我将原来的英文课程讲稿译成了中文。但是，由于种种原因，这两本书拖了四年多还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩，每本书都写得通俗易懂。中山大学马骏教授向格致出版社何元龙社长推荐了这套书，当格致出版社向我提出从这套丛书中精选

一批翻译，以飨中文读者时，我非常支持这个想法，因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种语言的精准把握能力，还要有对实质内容有较深的理解能力，而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容，只有语言能力是远远不能胜任的。在短短的一年时间里，我们组织了来自中国内地及港台地区的二十几位研究生参与了这项工程，他们目前大部分是香港科技大学的硕士和博士研究生，受过严格的社会科学统计方法的训练，也有来自美国等地对定量研究感兴趣的博士研究生。他们是：

香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛，硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊，应用社会经济研究中心研究员李俊秀；香港大学教育学院博士研究生洪岩璧；北京大学社会学系博士研究生李丁、赵亮员；中国人民大学人口学系讲师巫锡炜；中国台湾“中央”研究院社会学所助理研究员林宗弘；南京师范大学心理学系副教授陈陈；美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛；美国加州大学洛杉矶分校社会学系博士研究生宋曦。

关于每一位译者的学术背景，书中相关部分都有简单的介绍。尽管每本书因本身内容和译者的行文风格有所差异，校对也未免挂一漏万，术语的标准译法方面还有很大的改进空间，但所有的参与者都做了最大的努力，在繁忙的学习和研究之余，在不到一年的时间内，完成了三十五本书、超过百万字的翻译任务。李骏、叶华、张卓妮、贺光烨、宋曦、於嘉、郑冰岛和林宗弘除了承担自己的翻译任务之外，还在初稿校对方面付出了大量的劳动。香港科技大学霍英东南沙研究院的工作人员曾东林，协助我通读了全稿，在此我也致以诚挚的谢意。有些作者，如香港科技大学黄善国教授、美国约翰·霍普金斯大学郝令昕教授，也参与了审校工作。

由于所选每本书都有一篇序言，对相关方法的背景和应用作了很好的介绍，我们均予以保留，内容在此不再赘述。为了方便起见，我们将内容相似的书目集册出版，每册三至五本不等，共八册，它们分别是：《线性回归分析基础》、

《高级回归分析》、《广义线性模型》、《列表数据分析》、《纵贯数据分析》、《因果关系模型》、《社会科学中的数理基础及应用》和《数据分析方法五种》。所冠书名未必能精准涵盖其中的内容，读者可自行参阅每本书的序言或目录。

我们希望本丛书的出版，能为推动国内社会科学定量研究的扎实学风作出一点贡献。

吴晓刚

于香港九龙清水湾

目录

世代分析

序 3

第 1 章　引言 4

第 2 章　估计年龄、时期和世代效应的策略 15

第 3 章　用世代分析理解变革 38

第 4 章　数据要求与数据可用性 45

第 5 章　未来展望 54

注释 56

参考文献 57

译名对照表 60

纵贯研究

序 63

第 1 章　概述 64

第 2 章　纵贯研究的目的 67

第 3 章　纵贯数据收集设计 85

第 4 章　纵贯研究问题 94

第 5 章　纵贯分析 108
注释 135
参考文献 136
译名对照表 148

分析重复调查数据

序 153
第 1 章　导论 155
第 2 章　区分年龄、时期及世代效应 160
第 3 章　总趋势 166
第 4 章　分解总趋势 175
第 5 章　分解总变化的一般模型 190
第 6 章　发掘个人层次关系的变化 197
第 7 章　总结:分析社会变迁 219
注释 221
参考文献 223
译名对照表 228

多元时间序列模型

序 233

第 1 章　前言 235

第 2 章　对多元时间序列模型的介绍 237

第 3 章　基本的向量自回归模型 251

第 4 章　向量自回归分析范例 294

附录　多元时间序列分析的软件 320

注释 323

参考文献 325

译名对照表 329

世代分析

作者简介

诺瓦尔·D. 格伦(Norval D. Glenn)

任教于美国得克萨斯大学奥斯汀分校,是该校的阿什贝尔·史密斯社会学讲座教授和斯泰尔斯美国研究讲座教授。其主要研究兴趣是老龄化、生命历程和现代社会中的家庭关系。曾担任《当代社会学》和《家庭问题杂志》的主编,也曾是《美国社会学评论》、《民意季刊》、《婚姻与家庭杂志》、《人口学》和《社会科学研究》等杂志的编委会成员。

最近出版的作品涉及的主题主要包括向政策制定者和公众传播社会科学的研究发现、婚姻选择中的体制机制的变化以及婚姻中年龄和婚姻稳定之间的关系。

译者简介

於　嘉

2009 年毕业于中国人民大学,获法学学士学位,现为香港科技大学社会科学部在读硕士研究生。研究方向为社会分层、性别不平等。

序

许多社会科学学科都以“世代”研究为教学重心，例如人口统计学、流行病学、心理学、政治科学和社会学。格伦·埃尔德(Glen Elder)曾在其著作《大萧条中的孩子们——社会变迁对人生的影响》(*Children of Great Depression——Social Change in Life Experience*)一书中对“世代”这一概念进行了深入研究。该书讲述了一群孩子在成长过程中种种坎坷的经历，对他们而言，现代社会许多司空见惯的东西都是遥不可及、需要拼命争取才能拥有的。在本书的讨论中，我们以“迷惘的一代”为例，讲述他们怎样在战后努力向社会证明自己的能力。正是这些独特的经历造就了关于世代的研究。

一般而言，世代分析属于定量研究的范畴，研究者利用很多世代概念的估测方法对一些结果变量的效应进行研究。例如，在大萧条时期中，经历过战争的社会群体身上的某些行为特征。从这种意义上来讲，世代的成员都必须在特定时间段有同样的经历。另一方面，同一年出生的人(或特定时期出生的人群)经历了同样的社会变迁，或多或少都会表现出相似的性格特征。不同年龄的人有时也表现出这种群体特征。例如，同一年开始实习的毕业生们。所以在大多数情况下，“世代”指的是某一特定的出生人群。

20 世纪 70 年代出版的《大萧条中的孩子们——社会变迁对人生的影响》点燃了研究者的研究热情，他们开始寻找估测特定社会行为或社会现象以及估计世代效应的方法。这可并不简单，因为估测对象不仅仅包括世代(出生年份)的效应，还有年龄(出生后)以及特定时期(现有年份)的效应。众所周知，如果无法识别限制，那么 APC 线性模型和叠加模型也就无法被识别，因为年龄、时期和世代三个因素恰好构成一个线性函数式，其中年龄 = 时期 - 出生年，为一个恒等式。

诺瓦尔·D. 格伦发表的《世代分析》一书中提供了几种应对识别问题的新方法。作者在书中对世代分析进行了更深层次的讨论，内容涉及方法论基础——包括数据问题——并详细讲述了几种可供研究者选择的 APC 估测法。社会科学各个学科的研究者都会在此书中获得很多收获，那些研究年龄变化以及社会文化变迁的学者们将是最大的受益人。

廖福挺(Tim Futing Liao)

第 1 章 | 引言

世代分析并不是一项统计技术，而是检验数据时经常采用的策略。在过去的几十年中，随着可用数据量的不断增加，越来越多的社会科学研究者开始关注世代分析。虽然世代分析在与年龄增长及社会文化变革相关的重大问题中显示出有效性，但显然更多研究者是被其异常神秘的方法论难题所吸引的。毫无疑问，这种兴趣并非完全有益。许多针对世代分析的研究只是练兵，研究者花了许多精力去证明的问题其实在逻辑上不可行。这样的研究导致了许多过于自信，却很可能不正确的结论。

因此，我写这本书的主要目的，一方面是阻止研究者怀着这种错误的观念进行机械化的世代研究，另一方面是为了鼓励正确的研究方式。能更深入地了解世代分析的方法自然很好，但前提是研究者必须谨慎，且拥有足够的专业经验，这样才能正确破解其中的奥妙。

世代研究的目的

一般来说，研究者会在两种情况下使用世代分析，虽然这两种情况的目的都非常明确，但还是引发不少人对研究的目的进行一些区分。

在多数情况下，世代研究主要用于估测人类年龄变化所产生的效应，有时也估测婚姻产生的效应。显然，随机实验最适合用来估测诱因，但不能用于研究年龄变化产生的效应。我们无法将一定数量的个体随机分成实验组和对照组，无法观测实验组个体年龄变化的过程，并保证对照组不受与年龄变化相关的刺激因素的影响。因此，研究年龄变化效应的学生只能在一些非实验性方法

和准实验方法中进行选择，这些方法都有一定的局限性，而世代分析有独特的优点。我将在本章的下一节中讨论几种主要的方法。而本书的主要侧重点在于讨论与年龄变化相关的研究，但文中也经常出现世代用于分析其他情况的讨论。

世代分析的第二大用途是帮助研究者理解社会、文化及政治变革的根源和性质。这也是为什么世代分析在近几年中迅速成为关注焦点的原因之一。也许将来，这种用途会比前一种更加普遍。当然，除了世代法，肯定还有其他分析方法能够在社会和文化变迁的定量研究中发挥作用，但是世代研究所能达到的深度绝对是独一无二的，在这方面，没有任何一种方法能与之媲美。因此，我不会在本书中将世代研究和其他方法进行系统化的比较。尽管本书强调世代分析在研究年龄变化中的作用，但是我仍然会用一章的篇幅来讨论世代方法是怎样用于研究社会变迁的。

定义：世代研究与相关研究方法之间的比较

英文单词“cohort”原本用来形容士兵队伍，现在有时用来泛指一群拥有某些相同特质的人。而在本书及其他同期群分析的文献中，这个词汇被用来形容在特定时间段中经历过特定事件的人群。“世代”是社会学研究中最常见的同期群研究对象，所谓“世代”，就是指特定年份、某个 10 年间或其他时间段内出生的人群。如果没有限定词特指，那么文献中所说的同期群研究就是世代研究。在其他情况下，研究者用特定事件来特指同期群研究，比方说，曾经有研究者以退休组和毕业组为对象。如果同期群不是人类，那就必须增加一个限定词。同期群的种类很多，有以婚姻，也有以参加组织活动来定义同期群。虽然同期群的指代对象一般都是世代①，但也可以是婚姻、组织、课本、电影、车模，或其他任何在特定时间段内发生的事件。

---

① “Cohort analysis”有多种翻译，如队列分析、出生组分析、同期群分析、同辈群体分析、世代分析等。因为后文涉及的主要分析对象都是出生世代，而研究中最常见的应用也是出生世代，所以本书中统一译为“世代分析”。——译者注

很不幸的是,社会科学文献中居然曾出现"年龄世代"这样的定义,由于缺少事件因素,这种说法是不符合规则的。年龄当然是条件之一,而且是不断变化的条件,但是年龄绝对不是时间。如果研究中出现"年龄世代"这样的词汇,我们就很难确定其真正的含义——是指世代,还是简单的表示某一年龄段的人?

这两种含义是截然不同的。打个比方,所有出生在 20 世纪 60 年代的人被称为"60 世代"("60 后"),但是根据研究发生的时间,所有个体当时的年龄都是不同的。而只有 30 岁的人才能以年龄归为一类。也就是说,世代每往前退一年,年龄组都会发生变化。

在通常情况下,考虑到一个因变量需要两个时间点的信息才能被测量,所以只有两个或两个以上的世代进行对比的研究才能称为"世代分析"。不是所有以某一世代为对象的研究都是世代研究。比如从 2002 年 11 月开始,有研究者分别对青年群体、中年群体和老年群体的政治态度进行调查,并进行对比;也有人在 1995 年将某高校 1985 年所有毕业生的政治态度进行对比。这两项研究都不是真正的世代研究。前一项是横截面研究,数据从一个时间点或时间段中收集,后者是固定样本跟踪研究,测量同一群体的某些个性,从多个时间点收集数据。两种研究各有所长,都是非常有用的研究,但两者都无法满足绝大多数世代研究的要求,无法测量年龄增长所带来的影响,也就是年龄效应。

横截面数据也能显示年龄所导致的差别,但不能反映年龄效应,因为不同年龄的人(或其他研究对象)本身就属于不同的世代,其经历不同,产生的影响也不同。换句话说,这些不同就是所谓的"世代效应"。举个例子,20 世纪 20 年代出生的美国人在青少年时期经历了大萧条,而 10 年之后出生的人在他们的青少年时期见证了美国前所未有的经济增长。我们有理由相信,这两个世代因此拥有截然不同的政治态度和政治行为,一辈子都不会改变。在出生率较低的年代出生的人被称为"小型生育世代",他们一生中的经济生活、教育、事业和婚恋的机遇和大型生育世代的情况存在很大区别,而这些区别也影响着他们各自的态度和行为。因此,之所以不能用横截面研究来测量年龄效应或世代效应,是因为它会混淆研究结果。比方说,研究者无法弄清导致中年人和老年人之间观

念区别的真正原因，到底是因为老年人年长，其经历有别于中年人，还是因为各自的出生年代不同而导致的差异？

横截面研究还有一个局限，导致这种研究不适合在年长和年轻的存活个体间测量年龄效应，因为这会导致因变量与寿命之间的相关性。例如，美国21世纪初的全国调查显示，中年人饮用酒精饮料的比例高于老年人。这个发现本身并不能证明随着男性年龄增长，酒精饮料的饮用量会减少。这种差别当然可能部分是由世代效应导致，但也可能是一种组成效应，因为饮酒较多的人可能比较少饮酒或从不饮酒的人活得更短。大多数世代分析都使用从一系列截面研究中获得的数据，因此他们将不同死亡率的效应与年龄或世代效应混淆。

固定样本跟踪研究可以验证由不同死亡率造成的组成效应，同时它们也可以处理不断变老的世代中的个体层面和汇总层面的变化。但是，固定样本跟踪研究自身也不能为年龄效应提供具有说服力的证据，因为并不是所有随着个体年龄增长而出现的变化都是由于个体老龄化所产生的。尤其是在现代社会中，人们并不是在一个静止的社会中不断变老，而是处在一个不断变化的社会中，随着个体的变化，社会或文化变化的影响也作用于他们，这便造成了态度、行为、健康和情感的变化，并抵消了老龄化在静止社会中的效应。在跟踪数据研究中，这些时期效应与年龄效应相混淆。例如在美国，在20世纪70年代进入青年期的世代可能在20世纪80年代后期变得相对保守。我们无法判断这样的转变有多少是源自老龄化，因为整个社会也在向着同样的方向变化，那么就不能分辨世代间的趋势是不是源于时期的影响。因此，对20世纪70年代中期及10年后的高中生的政治态度的跟踪研究并不能为他们向成年人的转变提供任何有力的证据。

用固定样本跟踪来研究年龄效应的另一个局限在于，受访者可能受他们在调查中的参与情况的影响。如果这种效应发生在因变量上，那么这就是一个固定样本跟踪研究的条件习惯效应。在跟踪数据中，这一效应与年龄和时期效应会混淆。另一种阐释这个问题的方式即，如果存在固定样本跟踪研究的条件习惯效应，随着研究和调查（重复收集数据）次数的增加，一个原本能代表整体的样本将变得不再那么具有代表性。

固定样本跟踪研究的条件习惯效应可能发生在很多情况下。例如，如果因变量是根据一个标准测验测量出的能力或者技能，则可能存在一种特殊的固定样本条件习惯效应，称为“练习效应”，也就是说，随着固定样本跟踪研究一次又一次的进行，研究对象可能适应了测试而有更好的表现。同样，不断重复询问有关某一主题的尖锐问题，可能使受访者进行深思从而导致态度上的变化。或者，向受访者详细询问其工作、婚姻或者其他生活方面，可能会使他们渐渐对这类问题产生抗拒。受研究对象年龄增长影响的固定样本跟踪调查非常容易受固定样本条件习惯效应的影响，因为受访者可能习得了有效的方法来对他们在研究中的行为进行调节。

我们可以对固定样本追踪数据进行世代分析，但是很少有研究这样做，而且使用两个或两个以上截面研究的一般的世代分析可以避免固定样本调适效应。在这种称做“重复截面”的设计下，没有个体在多于一个时点接受了调查。相反，研究对来自每个世代的不同个体组成的样本在不同的时间进行了研究。

解释世代分析逻辑的一个启示性工具就是标准世代表，它通过将不同截面数据并排放置，显示出年龄与其他因变量之间的关系，而年龄的区间与进行数据调查的区间是一样的。表 1.1 正显示了一个这样的例子，它显示了美国人口普查数据中已婚妇女的比例。

表 1.1　根据年龄和年份分类的女性已婚的比例(美国)

年　龄	年份			
	1968	1978	1988	1998
25—34	87.4	76.6	67.3	67.3
35—44	87.1	82.1	76.3	72.1
45—54	82.4	80.5	76.2	70.8
55—64	67.7	70.4	70.7	67.8
65—74	46.5	48.3	53.3	54.8

资料来源：美国人口普查局进行的长期人口调查。百分比是根据美国统计局提供的数据计算得到(1969：表 37；1979：表 51；1990：表 49；1999：表 63)。

这个表中的每一列都是一个截面数据，而其中年龄、世代和一部分组成效应相互混淆。每一行都是处于同一年龄的 4 个不同的世代，其中，时期和世代效应相互混淆。除了 1998 年年龄在 25—34 岁之间和 1968 年年龄在 65—74 岁之间的世代，表中的每一个世代都可以跟踪至少 10 年，即表中从左上方至右下方对角线所显示的。在每一个世代对角线数据中，年龄和时期效应都相互混淆，而这个例子中的一小部分组成效应可以在中年后进行追踪。

因为数据来自截面研究，标准世代表中的每一列都受到效应混淆的影响，而每一个世代对角线的数据都与固定样本追踪研究一样会将年龄和时期混淆。然而，在世代表中，有许多列和许多世代对角线，这使得研究者希望能够通过统计方法分离出年龄、时期和世代效应。多年前，我将试图利用统计方法分离出世代数据中各种效应的方法称做“徒劳的追寻”(Glenn，1976)，除非在极少数的情况下所有的效应都线性无关。然而，并不是所有人都放弃了这种尝试。试图寻找一个能够一劳永逸地应用于估计效应的统计方法，是科学史上重复进行那些逻辑错误的尝试中最讽刺的例子之一。

识别问题

由于不可能利用统计方法分离出年龄、时期和世代效应，这就产生了识别问题，这一问题发生在当有 3 个或 3 个以上的自变量需要纳入研究，而每一个都是其他的线性函数的时候。换句话说，每一个自变量与其他自变量的复相关是唯一的——这也是共线性中可能出现的最极端的例子。如果所有的自变量都作为预测因子纳入回归方程或者类似的分析，则电脑程序不能运行。在其他变量都被控制的情况下，这些自变量的方差均为 0。例如，当我们有理由认为两个变量间的差别和两个变量本身都将影响因变量时，识别问题便发生了。一个例子即对婚姻质量的研究，丈夫的特征、妻子的特征和夫妻之间的差异都可能影响婚姻质量。另一个例子是社会流动性对幸福感的心理效应的研究，其中，出身的阶层、最终归属的阶层和两者的差别都可能影响到幸

福感。

在世代分析中,3 个相互关联的变量当然是年龄、时期和世代,其中每一个都是其他两个的完全线性函数。也就是说,根据其中两个变量的信息就可以推知第三个变量的信息。例如,如果我们知道一个在 1990 年进行访谈的调查对象当时是 20 岁,我们就会知道她的出生年月以及她的世代成员信息。或者,如果我们知道一个 1990 年进行访问的调查对象出生于 1970 年,则我们可以推知这个受访者当时的年龄为 20 岁。所以,如果时期和世代(或其他 3 个中的任意两个变量)都已经作为预测变量被纳入回归方程,那么再加入年龄就会造成冗余信息,使得电脑程序无法运行。如果有理由相信年龄、时期或者世代的任意一个没有效应,则剩余两个变量的效应很容易被估计,但是并没有直接的方法可以同时估计出这 3 个变量的效应。

识别问题的本质可以由表 1.2、表 1.3 和表 1.4 中的假设性数据所解释,其中对每个表格中因变量值的线性变异最简单的解释是,它们各自仅表现了年龄效应(表 1.2)、时期效应(表 1.3)和世代效应(表 1.4)(为了便于解释,我们假设因变量与寿命无关,因此也没有由不同的死亡率导致的组成效应)。然而,每个表中的数据同样适用于两个变量的解释,正如每个表底部的其他解释所显示的那样,而每个表中的变异模式都可能由年龄、时期和世代效应之间有限的组合所导致。显而易见,没有统计方法凭借其本身便可以从这些不同的效应组合中选择可以产生相同数据的组合。研究者的这种选择必须基于理论或者他们对现象来源的认识,而非基于数据本身。换句话说,选择必须基于菲利普·康威提出的“附加信息”(Philip Converse, 1976)。这种信息有时来自数据中记录的附加变量(除了年龄、时期、世代和因变量以外的变量),或者来源于其他途径。

研究者可能认为,尽管在逻辑上可以用 1 至 3 个变量对于世代数据线性模式的变异进行解释,但一个变量的解释往往是最有效的。然而事实并非如此。即便当变异的模式像表 1.2、表 1.3 和表 1.4 那么简单,2 个或 3 个变量的解释往往与理论或者其他来源的数据更加契合。此外,哪怕只是简单的线

性变异，也往往比表格中的假设数据复杂很多，而复杂的模式就需要复杂的解释。

表 1.2 单纯显示年龄效应的数据模式(抵消了时期和世代效应，或者年龄效应与相互抵消的时期和世代效应的混合)

年 龄	年 份					
	1950	1960	1970	1980	1990	2000
20—29	50	50	50	50	50	50
30—39	55	55	55	55	55	55
40—49	60	60	60	60	60	60
50—59	65	65	65	65	65	65
60—69	70	70	70	70	70	70
70—79	75	75	75	75	75	75

注：单元格中的数据是某个因变量的假设值。几个解释为：(1)每 10 年的衰老将导致因变量 5 点的增加；(2)每 10 年的变化对因变量产生 5 点的正向时期效应和 5 点的负向世代效应；(3)每 10 年的变化使因变量存在一部分年龄效应和相互抵消的时期和世代效应的混合。这些效应的有限组合可以生成表中所示的因变量的变异模式。

表 1.3 单纯显示时期效应的数据模式(抵消了年龄和世代效应，或者时期效应与相互抵消的年龄和世代效应的混合)

年 龄	年 份					
	1950	1960	1970	1980	1990	2000
20—29	30	35	40	45	50	55
30—39	30	35	40	45	50	55
40—49	30	35	40	45	50	55
50—59	30	35	40	45	50	55
60—69	30	35	40	45	50	55
70—79	30	35	40	45	50	55

注：单元格中的数据是某个因变量的假设值。几个解释为：(1)每 10 年的变化将导致因变量 5 点的增加；(2)每 10 年的时间变化对因变量有 5 点的正向年龄效应和 5 点的正向世代效应；(3)每 10 年的变化使因变量存在一部分时期效应和互相抵消的年龄和世代效应的混合。这些效应的有限组合可以生成表中所示的因变量的变异模式。

表 1.4 单纯显示世代效应的数据模式(抵消了年龄和时期效应,或者世代效应与相互抵消的年龄和时期效应的混合)

年龄	年份					
	1950	1960	1970	1980	1990	2000
20—29	50	55	60	65	70	75
30—39	45	50	55	60	65	70
40—49	40	45	50	55	60	65
50—59	35	40	45	50	55	60
60—69	30	35	40	45	50	55
70—79	25	30	35	40	45	50

注:单元格中的数据是某个因变量的假设值。几个解释为:(1)每一个到达了成年的世代将比之前的世代的因变量高出 5 点;(2)每 10 年的变化对因变量有 5 点的负向年龄效应和 5 点的正向时期效应;(3)每 10 年的变化使因变量存在一部分世代效应和互相抵消的年龄和时期效应的混合。这些效应的有限组合可以生成表中所示的因变量的变异模式。

例如,美国的年龄与工作满意度这个案例(见表 1.5)。工作满意度的整体水平在过去的几十年里相当稳定,且年龄与工作满意度之间有稳定的正向相关关系。对于这个稳定关系的最简单解释就是,它反映了年龄效应或者年龄效应与由于进出劳动力市场导致的组成效应的混合。

表 1.5 表示对工作"非常满意"[a] 的人数百分比[b]

年份	年龄					
	25—34	35—44	45—54	55—64	65—74	总计
1972—1976	44.4 (1475)	53.9 (1208)	51.8 (1358)	60.1 (936)	63.4 (313)	52.4 (5290)
1977—1980	41.3 (989)	49.0 (769)	52.2 (684)	61.1 (556)	60.6 (223)	50.2 (3221)
1982—1986	42.4 (1868)	49.1 (1384)	54.1 (1106)	57.2 (865)	62.4 (320)	52.9 (5543)
1987—1991	40.1 (1614)	46.5 (1553)	50.9 (1062)	52.2 (655)	57.5 (277)	49.4 (5161)
1993—1996	39.3 (1483)	44.6 (1563)	47.3 (1246)	55.9 (584)	68.1 (183)	45.9 (5059)

续表

年份	年龄					
	25—34	35—44	45—54	55—64	65—74	总计
1998—2002	46.3 (1275)	45.6 (1394)	49.1 (1149)	56.5 (539)	61.6 (189)	48.7 (4546)
总计	42.2 (8704)	47.8 (7871)	50.8 (6605)	57.3 (4135)	62.1 (1505)	48.9 (28820)

注：a. 其他回答为"一般满意"、"不太满意"和"非常满意"。
b. 通过加权近似获得了受访者的人数。权重是家庭中成年人的比例除以 GSS 家庭中的成年人平均数量。
资料来源：全国民意研究中心（NORC）于 1972—2002 年进行的年度美国综合社会调查（GSS）。

表 1.6 因变量的非线性变异可以解释为仅仅反映了年龄或者组成效应的数据模式

年龄	年份					
	1950	1960	1970	1980	1990	2000
20—29	50	50	50	50	50	50
30—39	52	52	52	52	52	52
40—49	62	62	62	62	62	62
50—59	62	62	62	62	62	62
60—69	50	50	50	50	50	50
70—79	45	45	45	45	45	45

注：单元格中的数据是某个因变量的假设值。

然而，整体水平的稳定性和工作满意度的年龄模式也迫使我们考虑这样的事实，即工作的条件在很多方面都有所变化（可能变得更好），而近些年连续进入职场的年轻人世代明显对于工作条件和回报有着更高的期望。因此，这个稳定的年龄模式似乎部分反映了相互抵消的时期和世代效应。年龄和工作满意度之间关系对线性的偏离，即每一年年龄的增长带来的工作满意度的增加在中年阶段比年轻和年长阶段较小，则显示了年龄或组成效应，这使得 3 个或 4 个变量的解释对这个例子更加适合。

表 1.2、表 1.3 和表 1.4 中因变量的变异都是线性的，但是年龄、时期和世代的非线性效应并没有像线性效应那样混淆。例如，表 1.6 中因变量指定为年

龄变异的非线性模式。对这个数据仅有的解释为，在没有组成效应存在的情况下，它们反映了非线性的年龄效应，因为年龄的非线性模式穿过了不同的世代和时期。如果时期的非线性模式穿过了不同的年龄和世代，或者世代的非线性模式穿过了不同的年龄和时期，则数据的解释也非常清楚。如果这个数据反映了超过一种效应，那么对效应的估计当然无法通过对列联表的简单观测来完成。但是，正如我将在下一章解释的，如果所有的效应都是非线性的，那么在统计上是可以依据一定的精度估计年龄、时期和世代效应。但是，因变量分年龄、时期和世代对线性变异的偏离并不意味着这些效应可以在统计上分离出来。如果变异的任何主要组成是线性的，那么对这些效应的统计分离是不可能的。

尽管识别问题涉及对数学定理的基本且广泛的理解以及除了世代分析之外的其他研究方法（例如，估计流动或者丈夫一妻子间不同的效应），但直到20世纪60年代，这个问题才被大部分社会科学家意识到。当时小休伯特·M.布拉洛克（Hubert M. Blalock, Jr.）发表了一系列关于这个问题的文章（Blalock, 1966、1967）。直到1973年才出现了一篇之后被广泛引用的文章讨论了世代分析中的识别问题（Mason, Mason, Winsborough & Poole, 1973）。这篇杰作的作者用简单的语言解释了涉及的逻辑和数学问题。但不幸的是，这篇文章对利用统计方法解决识别问题提出了不切实际的期望，因此，它也造成了试图分离世代数据中这些混合效应的一系列错误的尝试。其中一些将在下一章进行介绍。

第2章 | 估计年龄、时期和世代效应的策略

梅森、温斯伯勒和普尔等人的方法

在世代分析中用于估计年龄、时期和世代效应最常用的方法就是梅森等人在1973年的经典文章中所介绍的方法。这个方法和它的一些变形作为有力的方法一直被沿用下来。

梅森等人的方法的吸引力在于其简单易行。年龄、时期和世代都被记录为一系列虚拟变量，每一个虚拟变量代表5年或10年。当一系列变量作为预测变量被纳入回归，则其中的一个变量必须被忽略以使程序能够运行。当将年龄、时期和世代虚拟变量包含进来的时候，要使程序可以运行，我们就必须忽略一个额外的虚拟变量，它可以根据年龄、时期或者世代来构成。忽略这个引入分析的附加变量与忽略一系列变量中的两个变量的效果是一样的，并且这一假设往往与真实状况并不相悖，尤其当被忽略的两个虚拟变量代表的类别是相邻类别的时候。

在分析中同时加入年龄、时期和世代作为预测变量的方法当然不是唯一能够使程序运行的方法。有许多方法可以打破这些变量间的线性相关并使程序运行，例如对变量进行重新编码或者转换，即加以识别的限定。例如，年龄和世代可以作为连续变量而时期作为一系列虚拟变量加入回归。然而当我们这么处理的时候，线性相关的问题也仅仅在统计模型中被打破，但在现实世界中并没有被打破，因此，获得的估计的效应并不具有实际意义。

因此当我们决定使用梅森等人的方法时，这种选择并非随意的。许多研究者相信，由于相邻的年龄、时期或者世代的类别的效应是相同的这一假设并不违背

实际,因此选择任意两个相邻的类别就可以。但这是梅森等人从未提及的。

许多使用梅森等人的方法或其变形的研究者也忽略了文章中介绍的另一个重要部分,而这一部分内容由于出现在文章的脚注中,所以很难被注意到:

> 这些纯粹的效应(用于解释这一方法的假设效应)的形式是刻意设定为非线性的……我们将数据构造成这种形式的原因在于,完全的线性效应是难以被解释清楚的,而且在世代分析中,根据纯粹效应的变量和两个剩余变量估计的结果是一样的。(Mason et al., 1973: 248)[1]

换句话说,这种方法仅适用于估计非线性效应。

当效应是线性的时候,这种方法的弱点就在表 2.1 中显现了,其中年龄、时期和世代可根据基于不同简化假设的模型来估计出其效应。为了使分析更简单易懂,我构造了一个数据集,它符合每 10 年对因变量有 5 点的正向时期效应以及 5 点的负向世代效应这一假设,且在 1950 年时,20—29 岁之间的受访者的因变量值为 50。表 1.2 显示了构造后的数据结果。对于这个模拟的实验,我知道计算这些效应可以利用梅森等人的方法并观察其运行状况,因此我使用了 4 个具有不同简化假设(识别限定)的模型。表 2.1 是模型的估计结果。

表 2.1 利用梅森等人的年龄—时期—世代模型估计反映在表 1.2 中的效应(未标准化的回归系数)

变 量	模 型	1	2	3	4
常数项		50.0	50.0	25.0	28.7
	年 龄				
	20—29	a	a	0.0	a
	30—39	5.0	5.0	0.0	a
	40—49	10.0	10.0	a	3.6
	50—59	15.0	15.0	a	6.1
	60—69	20.0	20.0	0.0	8.6
	70—79	25.0	25.0	0.0	11.4
	时 期				
	1950	a	0.0	a	a

续表

变 量	模 型	1	2	3	4
	1960	0.0	0.0	5.0	a
	1970	0.0	0.0	10.0	3.6
	1980	0.0	a	15.0	6.1
	1990	0.0	a	20.0	8.6
	2000	0.0	0.0	25.0	11.4
	世 代（出生年份）				
	1871—1880	0.0	0.0	50.0	25.0
	1881—1890	0.0	0.0	45.0	23.9
	1891—1900	0.0	0.0	40.0	21.5
	1901—1910	0.0	0.0	35.0	19.1
	1911—1920	0.0	0.0	30.0	16.8
	1921—1930	0.0	0.0	25.0	14.0
	1931—1940	0.0	0.0	20.0	11.8
	1941—1950	0.0	0.0	15.0	9.1
	1951—1960	a	0.0	10.0	6.5
	1961—1970	a	0.0	5.0	3.9
	1971—1980	0.0	a	a	a

注：a. 参照类，其值为0。

除非相邻两个年龄层的效应是一样的（这样可以获得正确的估计），否则这个方法给出的结果大体上是不正确的。如果两年或者两个世代效应相同的假设是适用的，那么这个方法产生了一个单变量（完全的年龄效应）解决方法，同时，假设这一效应在3个变量解释中对两个年龄层和两年的结果是相同的。数据本身并不能提供应该使用哪个相同假设的线索，因为每一个假设对于现实的扭曲程度是相同的（当然，相比一个假设，两个假设引入了更多的扭曲）。显然，能够利用模型获得正确的估计取决于对哪一种解释的预先了解是正确的。这种方法并不能告诉研究者究竟是变量中哪一种组合的解决方式最能精确地代表现实。第1章中介绍的其他方法同样不能解决这个问题。但是，由于非线性的年龄、时期和世代效应彼此之间的混合与线性的不同，人们可能会认为，这些效应可以由梅森等人的方法来很好地估计。另一个模拟的实验可以提供相关

的证据。这一次，我假设非线性的纯粹年龄效应，其中，因变量在20—29岁群体中的值为50，且此后每10年的衰老具有的效应分别为+2、+10、0、-12和-5，直至70—79岁年龄段。这种模式的效应将产生表1.6中的数据。我再一次使用梅森等人的方法，用4种不同模型来估计反映在数据（表2.2）中的效应。对于模型1，其背后的假设为40—49岁的效应和50—59岁[①]的效应是一样的，而这是符合现实的，所以这个模型获得了准确的估计。对于模型2，其背后的假设为，任意两个世代的效应是相同的，而这也是正确的，因为没有世代效应，从而使得APC的所有估计都是正确的。对于模型3，其背后的假设为，任意两个时期的效应是相同的，同样，正确的简化假设使得对年龄、时期和世代效应的估计都是正确的。对于模型4，其背后的简化假设为，20—29岁和30—39岁的效应是一样的，但这并不正确，尽管误差并不大，但是可能导致与真实效应完全相反的估计。鉴于一个变量（纯粹年龄效应）的方法是正确的，这个模型显示了一个三变量模式的效应。

表2.2 利用梅森等人的年龄—时期—世代模型估计反映在表1.6中的效应（未标准化的回归系数）

变 量	模 型	1	2	3	4
常数项		62.0	50.0	50.0	60.0
	年 龄				
	20—29	-12.0	a	a	a
	30—39	-10.0	2.0	2.0	a
	40—49	a	12.0	12.0	8.0
	50—59	a	12.0	12.0	6.0
	60—69	-12.0	0.0	0.0	-8.0
	70—79	-17.0	-5.0	-5.0	-15.0
	时 期				
	1950	a	a	a	a
	1960	0.0	0.0	a	2.0
	1970	0.0	0.0	0.0	4.0
	1980	0.0	0.0	0.0	6.0

① 原书此处为30—39和40—49。应该是笔误。——译者注

续表

变　量	模　型	1	2	3	4
	1990	0.0	0.0	0.0	8.0
	2000	0.0	0.0	0.0	10.0
	世　代 （出生年份）				
	1871—1880	0.0	0.0	a	20.0
	1881—1890	0.0	0.0	0.0	18.0
	1891—1900	0.0	0.0	0.0	16.0
	1901—1910	0.0	0.0	0.0	14.0
	1911—1920	0.0	0.0	0.0	12.0
	1921—1930	0.0	0.0	0.0	10.0
	1931—1940	0.0	0.0	0.0	8.0
	1941—1950	0.0	0.0	0.0	6.0
	1951—1960	0.0	0.0	0.0	4.0
	1961—1970	0.0	a	0.0	2.0
	1971—1980	a	a	0.0	a

注：a. 参照类，其值为0。

从中获得的经验是显而易见的：哪怕效应的非线性模式简单地与表1.6中的一样，但只有当简化假设是正确的时候，梅森等人的方法才能够获得正确的估计。然而，一个研究者很少能够确信某一个简化假设是一点都不含有误差的。

当效应的非线性模式更复杂的时候，即包括两个或全部APC变量的非线性模式，梅森等人的方法或者其他的统计模型都不能充分解决问题。哪怕非线性效应彼此之间并不像线性模式中那样相互混淆，但它们之间仍然可能存在APC变量之间的交互。例如，可能非线性的年龄效应在统计分析中反映为时期和世代之间的交互，但并非所有的时期—世代交互都是年龄效应。表1.6中的非线性变异的简单模式使我们很容易判断出变异反映了年龄效应，但是更复杂的数据则无法通过简单的观察得出结论。

没有考虑到APC之间的交互是梅森等人的方法并不像很多研究者相信的那样有用的原因。所有关于这个方法的变形都基于年龄、时期和世代效应是叠

加的这一假设，也就是说，年龄效应对于所有的时期和世代都是相同的，时期效应对于所有的年龄和世代都是相同的，以及世代效应对所有的年龄和时期都是相同的。这个假设可以进行检验，但往往并不符合。当态度和行为的变化发生在成年人群体中时，变化往往在较年轻的人中更为突出。数据本身并不能分辨年龄—时期和世代—时期的交互，但是有理论表明，随着年龄的增大，成年人由于刺激而产生的变化逐渐变小（Glenn，1974、1980；Alwin，Cohen & Newcomb，1991）。态度的变化往往随着年龄的变化而不同，无论是更加激进还是更加保守，这也是我下一章即将讨论的问题。此外，一些年龄效应（尤其是那些最近发生的）与社会角色而非生物衰老相关的变化，也造成世代间的不同，例如对中老年人性生活活跃度的期望有所上升。世代效应也可能随着年龄而变化，例如一小部分世代群体在经济和职业上的优势随着他们年龄的增长而扩大，却又随着更年轻世代群体的挑战而慢慢减小。

通过对表 1.1 中女性结婚率的数据进行检验，可以揭示出叠加的 APC 模型的不完善之处以及利用附加信息解释世代数据的需求。这个数据中存在大量的交互。尽管结婚率在所有年代都随着年龄负向变化，但用年龄对 1968 年和 1998 年结婚率进行的回归结果显示，其系数从－1.012 变为－2.93。此外，虽然在 1968 年结婚率最高的是在 25—34 岁年龄段，但在其后的时期，结婚率最高的年龄段变为 35—44 岁。同时，用年份对结婚率进行回归，其回归系数从 25—34 岁年龄段的－0.696 变为 65—74 岁年龄段的 0.299。1968 年处于 25—34 岁年龄段的世代在迈向 35—44 岁年龄段时，经历了结婚率的下降，但是经过同样的衰老期，1978 年处于 25—34 岁年龄段的世代并没有经历结婚率的下降，而 1988 年处于 25—34 岁年龄段的世代则经历了结婚率的上升。

对这些强调交互的效应进行统计建模是困难的，哪怕识别问题并不会产生干扰。然而，对于熟悉美国近些年来婚姻、离婚和寿命的家庭人口学者来说，这种数据并非十分神秘。相关的趋势为：(1)从 20 世纪 70 年代末至 20 世纪 80 年代，典型的初婚年龄有显著的上升；(2)从 20 世纪 60 年代中期至 20 世纪 80 年代，离婚率有显著的上升，而持久的婚姻仅占很小一部分；(3)中老年死亡率从 20 世纪 80 年代之后又持续下降；(4)进入较老年龄段的世代具有最高的终身结

婚率。如果不了解这些趋势,对表 1.1 中的数据进行统计处理则并不能带来对数据复杂模式的深入理解。

中村的贝叶斯方法

梅森等人的世代分析方法只有在简化假设(识别限定)是胡乱被选择的情况下才可以随意地作为一种常规方法来运用,但这是一种对方法的滥用。然而,其他世代分析方法则希望建立一套程序化的应用,也就是说,无论理论或者附加信息预测出怎样的年龄、时期和世代效应,这种方法都可以以相同的方式进行应用。

这些方法中最著名的可能是来自日本统计学家中村的研究(T. Nakamura, 1982、1986),并由佐佐木和铃木介绍给美国社会科学家(Sasaki & Suzuki, 1987)。这个方法不变的简化假设就是连续的参数是逐步变化的。更具体地说,这个假设就是在年龄、时期和世代中,相邻的比分散的更为接近。佐佐木和铃木给出了进一步的解释:

> 对于一个社会现象,较短间隔内进行调查产生的百分比变化比较长间隔内进行调查产生的百分比变化要小。例如对宗教的热忱度,20 岁年龄段和 30 岁年龄段群体之间的差距大致上比 20 岁年龄段和 40 岁年龄段之间的差距小。(Sasaki & Suzuki, 1987:764)

佐佐木和铃木承认,连续的参数是逐步变化的这一假设并非永远正确。研究者只需要观察 APC 变量中任意两者与因变量之间的关系是否是单调即可。如果确实如此,或者只有很少的违背之处,则这个假设基本上是正确的。

如果佐佐木和铃木对他们观点的阐述是有效的,那么这个方法的确是有用的,因为逐步变化这个假设可以实证地进行检验。佐佐木和铃木将他们的方法应用到与表 1.4 非常类似的世代数据中(Sasaki & Suzuki, 1989),所有由年龄和时期导致的变异不仅是单调的,而且是完全线性的。尽管有无数种可能的解

释且其中一些对于特定的因变量是合理的，但是他们总结道，数据只反映了世代效应。这显示出这个方法往往选择所有可能的解释中最简单的，至少是像这个方法所限定的简约性一样。这种解释可能通常是正确的，但仅仅是“通常”，这对于建议研究者使用这个方法且接受这个估计是正确的理由并不充分。同样，相比于考虑了更多信息而并不十分有力的方法得出的结论，中村的估计方法并不更加准确。

显然，对中村的方法应该持怀疑的态度，但是如果估计是根据理论和附加信息或者是对数据模式的评估信息而获得的，那么便可能是有效的。例如表 2.3 中，佐佐木和铃木对荷兰关于无宗教归属人士的数据进行的分析(Sasaki & Suzuki, 1987)。在一个独立的表中，作者指出，在所有的荷兰成年人口中，无宗教归属人士从 1899 年的 1.8%单调地增长至 1969 年的 23.0%。然而从表 2.3 的数据中可以看出，时期效应对于无宗教归属有很强的影响。当这样的时期影响持续时，那么便会出现一个与全部人口中变化方向一样的世代间的变化趋势。因此，任何检视表 2.3 的分析者都能发现，该表显示出了数据中较大的正向时期和世代效应。尽管非宗教归属在表中每个世代内都经历了超过一个时期的上升，但并没有其他来源的理论或证据显示，随着年龄的增长宗教归属感会下降。因此，世代间的变化很大程度上应该是时期效应。

表 2.3　荷兰无宗教信仰人口百分比(根据年龄和年份)

年份	年龄					
	20—29	30—39	40—49	50—59	60—69	70—79
1899	2.4	2.2	1.7	1.2	0.9	0.6
1909	5.8	5.6	4.2	3.2	2.4	1.6
1919	8.4	8.9	7.1	5.3	3.9	2.7
1929	15.0	16.2	14.1	11.0	8.1	5.8
1939	16.9	18.3	17.0	14.6	11.2	8.0
1949	18.0	19.5	19.0	17.5	14.3	10.4
1959	18.3	20.2	19.8	19.2	17.8	14.2
1969	24.7	23.0	23.3	23.1	21.4	19.2

资料来源：佐佐木和铃木书中的表 4(Sasaki & Suzuki, 1987)。

表 2.4 中村的贝叶斯方法和一般最小二乘法估计得出的荷兰非宗教归属的年龄、时期和世代效应

	中村的贝叶斯估计	一般最小二乘回归
年龄		
20—29	—0.715	a
30—39	0.0641	a
40—49	0.0615	a
50—59	0.0473	a
60—69	—0.0063	a
70—79	—0.0950	a
时期(年份)		
1899	—1.2301	a
1909	—0.5361	1.352
1919	—0.2775	2.269
1929	0.2625	6.385
1939	0.3305	7.406
1949	0.3985	8.042
1959	0.4355	8.732
1969	0.6177	11.835
世代(出生年份)		
1820—1829	—1.2967	—5.686
1830—1839	—1.0314	—7.712
1840—1849	—0.7699	—5.393
1850—1859	—0.4908	—5.137
1860—1869	—0.2230	—4.208
1870—1879	0.0447	—2.578
1880—1889	0.2913	a
1890—1899	0.4545	2.286
1900—1909	0.5195	3.814
1910—1919	0.5575	4.536
1920—1929	0.5754	4.678
1930—1939	0.5539	4.081
1940—1949	0.7742	6.579

资料来源:格伦书中的表 3(Glenn, 1989)。

表 2.4 显示了中村对数据中效应的估计，且估计结果精确地反映了理论、附加信息和对数据的非正式评估所指示的效应模式。在这个例子中，这个方法显然非常实用。另一方面，研究者也可以通过简单的假设来估计不存在年龄效应的情况而获得类似的结果，将时期和世代都转换为虚拟变量，并使用最小二乘回归来估计时期和世代对非宗教归属的效应（参见表 2.4 中最后一列）。从这个例子中我们可以看出，复杂的中村方法并不比简单的回归分析有效，尽管在其他的例子中也许是这样。

中村的贝叶斯方法并没有被广泛地使用（至少在美国如此）。佐佐木和铃木认为，这种方法为世代分析中的识别问题提供了一个全面的解决方法，这一观点已经大致上被否定了。然而，这个方法有时可以对效应提供可信的估计，它也尽可能努力地成为分析世代的全面方法。因此，尽管中村方法并没有成功，但仍有许多后续的研究希望提供一个对世代分析普遍适用的统计方法。

后续的探索

这本书并不想过多地引用前人的著作，因此我也不能总结它们的内容或者介绍其作者。但至少有一些作者将其新的方法称做“解决 APC 难题的全能方法”以招徕读者，他们宣称，当数据为短时期多次收集获得的、可比较的重复截面数据时，应用这个方法将获得有效的估计。当这些方法利用已经发表的文章进行解释时，我们并不能要求其分析考虑到理论和附加信息以及在 APC 变量线性模式中因变量方差的含混性。换句话说，至少有一些方法的鼓吹者号称他们可以解决逻辑上不可能的问题。

当然，如果“通常”能够获得接近正确的估计，研究者能够使用理论和附加信息来评价估计的可信度，且他们对所获得的结论持探索性态度，那么这种方法也许可以被证明是有效的。

我对那些想要使用这种方法的研究者的建议是，不妨等到它已经被不同情境下模拟的实验状况完全地检验后再使用。或者，研究者可以自己进行这类模

拟。当模拟完毕后，研究者应该牢记，几个对效应成功估计的例子并不能够证明方法是普遍适用的。此外，在模拟实验中，也应该将复杂的效应混合情况考虑进去。

年龄—时期—世代—特征模型

对世代效应研究的最新进展是年龄—时期—世代—特征（APCC）模型（O'Brein，1989、2000；O'Brein，Stockard & Isaacson，1999）。这些模型包括一个或一个以上的“世代特征”，例如世代规模或者对随世代群体而变化的家庭结构的测量。尽管不能控制年龄和时期而单独让世代变化，但是当世代的特征与世代的相关程度不太高的时候，是可以控制年龄和时期而让世代的特征变化的。如果研究的目标仅仅是为了估计选择出来的世代特征的效应，而非年龄或时期的效应，并且世代特征与世代之间并没有强烈的线性关系，那么这些模型就可以被有效地应用。但是需要记住的是，这些模型并不是真正的 APC 模型（尽管有时它们也用来表示 APC），且它们并不能够解决年龄—时期—世代这个难题。

某个因变量的世代效应不可能仅仅是因为模型中包含的世代特征所导致的，而模型估计中的世代效应也与年龄和时期效应混淆不清，这样做可能导致模型估计出的时期与年龄效应没有意义。此外，如果世代特征与世代的线性相关较严重时，世代特征的效应也容易与时期和年龄效应相混淆。幸运的是，在最近的美国数据中，世代规模与世代群体的出生日期线性相关较低，因此利用 APCC 模型估计世代规模的效应也会较为准确。然而一些其他的世代特征，例如婚外所生的婴儿比例和在未成年期经历了父母离婚的世代成员的比例，则与世代有单调的线性关系（尽管并非完全线性）。这些特征都被应用在 APCC 模型中，但是它们与世代之间严重的线性关系会导致对其效应的估计不那么准确。[2]

当然，研究者可以同时构想年龄—特征—时期—世代模型和年龄—时期—特征—世代模型。尽管我没有看到此类方法的专有术语，但是在某些研究中，

时期、世代和一些特征与年龄是具有一定程度相关的，且它们同时被当做自变量，此外，还有些研究中使用了可以被称做“年龄—时期—特征—世代模型”的方法。与APC模型一样，这些模型同样具有长处与缺点，因而需要尽量避免自变量与忽略的APC变量具有的很强的线性关系，而这样也导致对包含APC变量的估计没有意义。

对APC效应估算的非正式方法

不能通过统计上的APC模型对年龄、时期和世代效应作出准确估计这一事实并不意味着我们要放弃对这些效应的区分。一旦不能精准和绝对确信地分离这些效应，那么利用理论、附加信息、常识以及多种统计分析来进行判断就是合理的。但是并没有公式或者手册适用于区分所有情况下的这些效应。在少数情况下，只需对交互数据进行目测即可，这时候我们会发现，因变量的变化与表1.6中的情形非常相似。同样，对于自变量在世代间变化的非线性模式在不同年龄和时期是一致的情况下，或者不同时期间变化的非线性模式在不同年龄和世代是一致的情况下，对此都只有一种可能的解释。但是对这些效应作出合理的判断往往是困难的，且需要研究者具有巧妙的心思、深谋远虑和良好的判断。

为了使年龄、时期和世代效应的估计获得足够确信的结论，我们往往需要以很多种方法来审视数据，并引入很多种附加信息(参见 Abramson & Englehar，1995；Alwin，1991；Converse，1976；Glenn，1994、1998)。而从简单到复杂的各种统计方法，也可能起到一定的作用。

一种简单而有效的方法就是对介入年龄、时期或者世代和因变量之间的变量进行统计控制。我们以美国为例来解释这个问题，在农村长大的人口比例随着不同的世代群体的变化而有很大的改变，那么，通过控制出生人口所在社区的规模来移除任何由年龄导致的自变量变异，则可能是世代效应而非年龄效应。例如，同胞数量或者父母是否离异这样的未成年时期的背景变量则可能同时介入世代与一些因变量之间。教育程度是另一个与世代相关的变

量，但是对控制它所获得结果的解释必须十分小心，因为教育对于不同的世代自然会有不同的效应（Alwin，1991；Glenn，1994）。例如，婚姻状况、父母状况、就业状况以及生理健康都可能介入年龄与各种因变量之间；富裕程度、失业率和政治大环境都是与时期相关的变量，而在世代分析中对其进行控制也较为有效。

由于非正式的世代分析方法必须根据所面对的特殊问题进行应用，我并不能对所有可以被应用的有效技术一一进行介绍和解释。但是，这里我举一个以个人幸福水平为因变量的例子来解释一些有用的策略。

实例：一项有关个人幸福感的世代研究

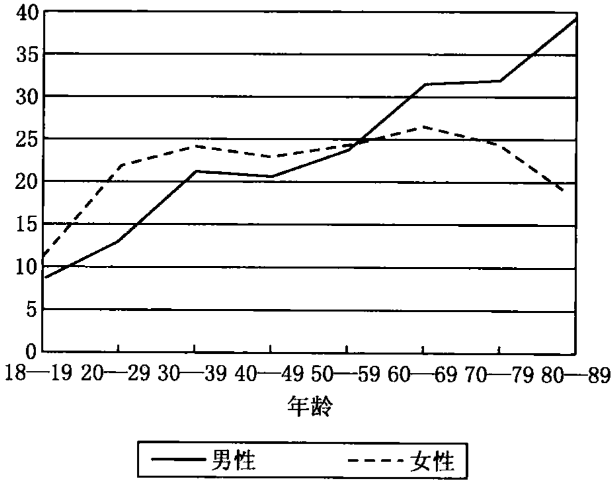
这个研究的目的在于尝试性地回答这些问题：第一，在美国社会中，成年人年龄的增长对个人幸福感有何影响？第二，这些影响在男性和女性间有什么差别？

这个研究的数据来源是芝加哥大学的国家民意研究中心 1972 年至 2002 年间进行的美国综合社会调查（GSS），这个调查从 1972 年至 1978 年每年进行一次，1980 年进行一次，1982 年至 1991 年每年进行一次，1993 年、1994 年、1996 年、1998 年、2000 年和 2002 年也各进行一次。在每次的调查中，受访者都会被问到如下问题：

> 综合来看，你最近的感受是非常幸福、一般幸福还是不怎么幸福？

在研究中，这个定序的三点度量往往被作为定距变量进行处理，这里我进行的回归分析也同样如此。然而，我使用了另外两个测量对汇总层次进行分析：(1)幸福指数，即用选择“非常幸福”的受访者比例减去选择“不怎么幸福”的受访者的比例；(2)对于选择“非常幸福”与选择其他的受访者进行一个简单的二分。

把从1972年至2002年的数据合并在一起所得的图2.1显示了幸福指数(HI)与年龄的关系(男女分别计算)。



注:a. 选择"非常幸福"的受访者比例减去选择"不怎么幸福"的受访者的比例。
资料来源:美国综合社会调查1972年至2002年的合并数据(Davis et al., 2002)。

图2.1 根据年龄和性别绘制的幸福指数[a]

对于男性,年龄与幸福指数的关系是单调递增的;对于女性,幸福指数值在中年的时候是最高的,在更高年龄时降低,而在青年时期是最低的。在50岁以前,相对于男性,女性有较高的幸福指数;但是50岁之后,这个关系则正好反了过来。

正如读者所知,幸福指数与年龄的关系也许被年龄、世代和删减(组成的)效应所混淆。对这个数据的第一反应的解释可能是男性和女性的幸福感是如何随着年龄而变化的,但是也有另外的解释。例如,男性幸福指数与年龄的单调正相关也由于每一个成年的世代群体都比之前的世代群体更加不幸福。此外,男性与女性的差别与年龄之间的关系也许是由于这些年轻女性在美国社会所获得的进步导致的,这也造成了与较老的女性出生群体相比,较年轻的女性出生群体与男性之间的差距较小。

表 2.5 根据年龄、时期和性别的幸福指数[a](美国,括号中为人数)[b]

年龄	1972—1982		1983—1992		1993—2002	
	男性	女性	男性	女性	男性	女性
18—19	4.4 (250)	9.0 (211)	11.9 (168)	12.6 (151)	11.2 (143)	13.4 (134)
20—29	8.8 (1059)	21.7 (1195)	17.7 (1447)	23.4 (1679)	13.6 (1261)	19.5 (1364)
30—39	19.3 (1218)	26.9 (1495)	21.9 (1343)	22.0 (1680)	22.2 (1416)	23.5 (1666)
40—49	19.6 (1158)	26.2 (1366)	22.1 (1074)	23.6 (1323)	19.8 (1348)	19.5 (1617)
50—59	23.4 (1183)	25.8 (1319)	23.3 (803)	20.1 (935)	25.2 (917)	25.8 (1078)
60—69	28.6 (829)	25.6 (906)	32.7 (707)	27.9 (922)	34.4 (591)	25.8 (704)
70—79	30.5 (471)	25.6 (510)	30.7 (407)	26.1 (583)	34.7 (427)	21.6 (555)
80—89	33.1 (118)	18.3 (147)	29.7 (128)	24.2 (223)	30.1 (146)	11.9 (252)
总计	19.1 (6958)	24.5 (7869)	22.7 (6069)	23.3 (7496)	22.3 (6249)	21.7 (7370)

注:a. 选择"非常幸福"的受访者比例减去选择"不怎么幸福"的受访者的比例。
b. 受访者的人数是根据权重近似获得的。数据的权重是家庭中成年人的比例除以 GSS 家庭中的成年人平均数量。

资料来源:美国综合社会调查 1972 年至 2002 年的合并数据(Davis et al., 2002)。

当合并的截面数据包括了几年甚至几十年时,我们必须观察获得的关系是否在这段时间内是稳定的。如果不是,那么其中的变化则可以显示出数据应该进行怎样的解释。对数据的目测观察并不能发现年龄与幸福指数之间的变化模式(无论男性还是女性),这一事实也同样被分性别和时期、根据年龄对于(幸福感)3 点度量的回归结果所验证。男性在 3 个时期的非标准化回归系数分别为 0.004、0.003、0.004,且每个系数的双尾检验在 0.01 水平上均显著。而类似的女性系数则为 0.001、0.001、0.000,但没有一个在统计上是显著的。因此,对合并数据进行解释时,不考虑年龄与幸福指数之间的关系在 1972 年至

2002 年间的变化是合理的。

表 2.6 美国人分时期、性别选择“非常幸福”[a] 人数的百分比(括号中的为人数)[b]

	男性	女性	男女差异
时期			
所有年龄			
1972—1982	33.2 (6973)	36.6 (7905)	−4.4***
1983—1992	32.8 (6084)	33.6 (7523)	−0.8
1993—2002	32.4 (6256)	32.7 (7388)	−0.3
变化			4.1***
年龄(18—34 岁)			
1972—1982	24.7 (2641)	34.8 (2935)	−10.1***
1983—1992	28.3 (2286)	31.6 (2712)	−3.3*
1993—2002	27.6 (2113)	31.4 (2328)	−3.8**
变化			6.3***

注:a. 其他的受访者选择的是“一般幸福”和“不太幸福”。
b. 受访者的人数是根据权重近似获得的。数据的权重是家庭中成年人的比例除以 GSS 家庭中的成年人平均数量。
* 表示 $p < 0.05$ (双尾检验); ** 表示 $p < 0.01$ (双尾检验); *** 表示 $p < 0.001$ (双尾检验)。
资料来源:美国综合社会调查 1972 年至 2002 年的合并数据(Davis et al., 2002)。

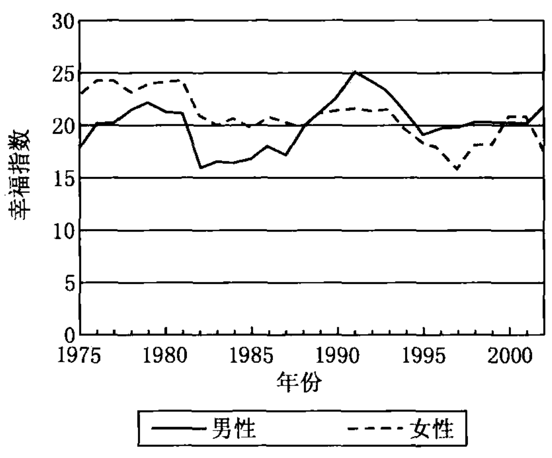
然而,1972 年至 2002 年间的变化是很重要的,且必须考虑进来。例如,如果利用世代效应来解释男性的幸福指数与年龄之间的正相关是正确的,那么男性整体的幸福程度应该在 1972 年至 2002 年间呈下降趋势,且在最年轻的成年人处下降最快。尽管不能完全拒绝此解释,但是表 2.6 和表 2.7 的数据对此解释显示出很强的质疑。这些数据显示了男性的平均幸福程度在上升,尤其在 18—34 岁的年龄阶段。同样,如果女性在较低年龄具有较高幸福水平和男性在较高年龄具有较高幸福水平完全或大部分是由于世代效应而

导致的，那么女性应该在 1972 年至 2002 年间不断逼近男性，尤其在年轻成年人中。然而，图 2.2 和表 2.6、表 2.7 显示出了相反的趋势。男性不断逼近女性，且与整体相比，这种变化在年轻成年人间更加明显。这种解释是基于没有严重的时期效应抵消了世代效应这一假设之上。这是一个合理的假设，因为时期与世代对整体趋势的影响效应往往（并非始终）相互加强，而非彼此抵消。

表 2.7 对美国人幸福度[a] 根据年（1972—2002 年）和性别回归的结果

	男性	女性	男女差异
所有年龄	0.002***	−0.001	0.003***
年龄（18—34 岁）	0.003***	−0.001	0.004**

注：a. 幸福水平是根据 3 点进行度量的，且处理为定距变量。
** 表示 $p < 0.01$（双尾）；*** 表示 $p < 0.001$（双尾）。
资料来源：美国综合社会调查 1972 年至 2002 年的合并数据（Davis et al.，2002）。

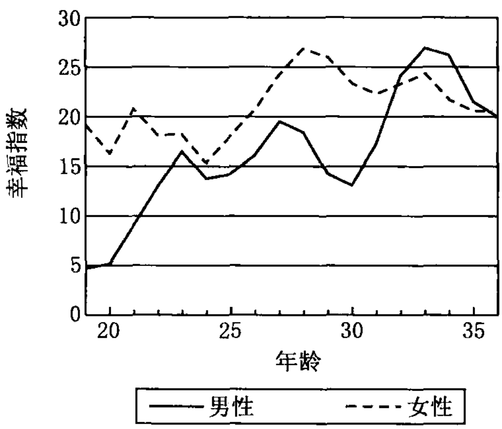


注：a. 选择“非常幸福”的受访者比例减去选择“不怎么幸福”的受访者的比例。
b. 所指年份为 4 年一轮的最后一年。
资料来源：美国综合社会调查 1972 年至 2002 年的合并数据（Davis et al.，2002）。

图 2.2 根据时期和性别的幸福指数[a]（4 年一轮的均值[b]，美国）

图 2.1 显示出的幸福指数与年龄之间的关系反映了年龄效应，这也可以由反映幸福水平随着世代群体年龄在 1972 年至 2002 年间增长趋势的数据加以补

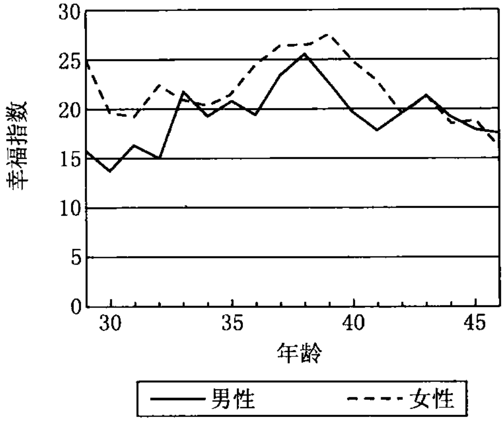
充说明。追踪世代群体间的趋势往往通过选择一个较宽的出生年份区间的出生群体(例如一个 5 年或 10 年的出生群体),并从数据中的第一年到最后一年对他们进行追踪。这个步骤的一个缺点在于,对世代中的年轻成员与较老成员进行的追踪并非经过同样的年龄。例如,我们选择 1972 年年龄为 20—29 岁的世代,并追踪他们至 2002 年,最老的成员是从 29 岁追踪至 59 岁,而最年轻的成员则是从 20 岁追踪至 50 岁。对某些研究而言,这个步骤已经足够了,但是当我们想要更清楚地了解在某个特殊的年龄发生了什么时,则需要对每个 5 年或 10 年的世代在同样的年龄进行追踪。当然,在进行此步骤时,不同世代的成员是在不同的时期进行追踪的。如果数据被限制为能够包含所有世代成员的时间点(这是为了使得样本规模足够大以获得可靠的估计),则时间的跨度可能变小。在 1972—2002 年的 GSS 数据中,第二种方法只能追踪 20 年的世代,而非 30 年。



注:a. 选择"非常幸福"的受访者比例减去选择"不怎么幸福"的受访者的比例。
b. 所指年龄为 3 年一轮中的中间年龄。
资料来源:美国综合社会调查 1972 年至 2002 年的合并数据(Davis et al., 2002)。

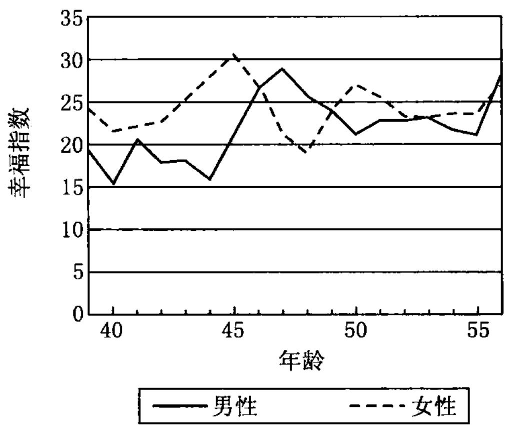
图 2.3 1955—1964 年世代的幸福指数[a](年龄从 18 岁至 37 岁,3 年一轮的平均值)[b]

在图 2.3 至图 2.7 中,我们显示了幸福指数在世代内部的趋势,共 5 个 10 年世代群体,从 1915—1924 年世代到 1955—1964 年世代,综合起来从 18 岁追踪至 77 岁。如果对图 2.1 中数据的年龄效应解释是正确的,那么在每个世代中,男性应该经历幸福指数的上升,且男性的趋势应该比女性的更加令人



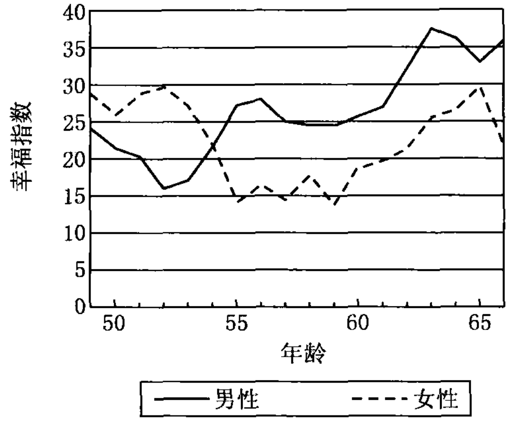
注：a. 选择“非常幸福”的受访者比例减去选择“不怎么幸福”的受访者的比例。
b. 所指年龄为3年一轮中的中间年龄。
资料来源：美国综合社会调查1972年至2002年的合并数据(Davis et al.，2002)。

图2.4 1945—1954年世代的幸福指数[a]（年龄从28岁至47岁，3年一轮的平均值）[b]



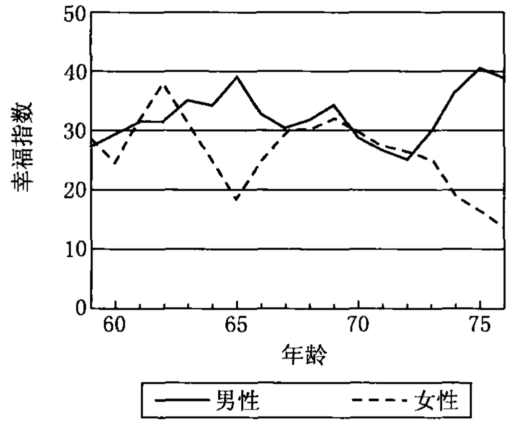
注：a. 选择“非常幸福”的受访者比例减去选择“不怎么幸福”的受访者的比例。
b. 所指年龄为3年一轮中的中间年龄。
资料来源：美国综合社会调查1972年至2002年的合并数据(Davis et al.，2002)。

图2.5 1935—1944年世代的幸福指数[a]（年龄从38岁至57岁，3年一轮的平均值）[b]



注:a. 选择"非常幸福"的受访者比例减去选择"不怎么幸福"的受访者的比例。
b. 所指年龄为3年一轮中的中间年龄。
资料来源:美国综合社会调查1972年至2002年的合并数据(Davis et al.,2002)。

图2.6 1925—1934年世代的幸福指数[a](年龄从48岁至67岁,3年一轮的平均值)[b]



注:a. 选择"非常幸福"的受访者比例减去选择"不怎么幸福"的受访者的比例。
b. 所指年龄为3年一轮中的中间年龄。
资料来源:美国综合社会调查1972年至2002年的合并数据(Davis et al.,2002)。

图2.7 1915—1924年世代的幸福指数[a](年龄从58岁至77岁,3年一轮的平均值)[b]

满意。图2.3至图2.7和表2.8也大致符合这一预测。事实上,预测出的变化比图2.1中显示的变化将导致产生年龄与幸福指数之间关系的更大变化。截面数据中男性幸福指数(3点度量)根据年龄的回归系数(非标准化)为0.002(图2.7),根据世代内趋势数据获得相应的每年的变化为0.006(世代的均值根

据表 2.8 中的数据计算获得)。此外,截面数据中男性与女性回归系数之间的差别为 0.003(表 2.7),但是对男女之间幸福指数预测的世代内每年均值的变化为 0.008(根据表 2.8 中的数据计算获得)。

表 2.8 20 年后对美国 5 个 10 年一组的出生群体的幸福感根据年龄分性别进行回归(非标准化)

世代 (年龄段)	男性	女性	男女差异
1915—1924 (58—77)	0.004	−0.006*	0.010*
1925—1934 (48—67)	0.009**	−0.004	0.013**
1935—1944 (38—57)	0.005	0.000	0.005
1945—1954 (28—47)	0.002	−0.003	0.005
1955—1964 (18—37)	0.009***	0.003	0.006*

注:a. 幸福水平是根据 3 点进行度量的,且处理为定距变量。
* 表示 $p < 0.05$ (双尾); ** 表示 $p < 0.01$ (双尾); *** 表示 $p < 0.001$ (双尾)。
资料来源:美国综合社会调查 1972 年至 2002 年的合并数据(Davis et al., 2002)。

对这个数据最合理的解释为,世代内的趋势是由年龄和时期效应混合导致的,时期效应明显是由时期的影响带来的,这在表 2.6 与表 2.7 中的年轻人的趋势中可以显示出来。换句话说,最近的时期效应使得男性的幸福指数相对女性升高,也使得逐渐老化的世代中的变化增添了同向的年龄效应。

表 2.6 与表 2.7 中显示的年轻人中的趋势表明了时期影响,但是它们同时也反映了世代效应。这些趋势在何种程度上可以被概念化为时期效应或者世代效应并不清楚,但是承认它们至少一部分是世代效应,这对于解释年龄和幸福感的截面数据是十分必要的。例如,任何世代内幸福程度升高的趋势将会降低由于年龄正向效应导致的截面的年龄差异。这意味着年龄对男性幸福感的正向效应往往比图 2.1 中截面数据所示的大。

美国在最近几十年来,男性的幸福感明显地随着年龄的增长而单调上升,

而且由于他们变老，正向的年龄效应也比数据显示的大，正如图 2.1 中所表现的那样。女性的幸福感符合同样的单调趋势，因为随着她们年龄的增长，与男性几乎所有的年龄水平相比，女性的年龄效应较低，且在年纪较长时呈负向。

当世代分析的范围中包含较大年龄时，则应该考虑由于不同的死亡率而导致对因变量产生的组成效应。例如在这个研究中，不幸福的人可能比幸福的人死得更早，而这样就造成了数据上，男性从 70 岁到 80 岁之间幸福感的大幅上升。同样，这种组成效应也可能发生在男性较年轻的阶段，并且也可能使得女性在较老时幸福感下降得没有那么严重。

并没有明确的证据能对这种状况发生的可能性进行评估。老年人所报告的幸福感与他们的长寿程度相关（Palmore & Jeffers, 1971），但这可能仅仅是由于人们在去世之前往往经历了健康程度的下降，而较差的自我感知健康程度与报告的幸福程度相关。如果是这样的话，那么死亡率对汇总层面的幸福感的正向效应则可能很大程度或者完全地被去世前健康程度下降的负向效应所抵消。在这里，我们需要但是难以获得的信息是有关幸福感与寿命在一生中的关联（随着年龄增长而产生的自我感知健康程度下降对男性的幸福感产生了明显的正向年龄效应）。

由此看来，我们认为死亡率差异并不是本研究中显示的、年龄和世代模式的原因。然而，如果很难找到对于这一发现其他的原因，这对于男性从 70 岁到 80 岁之间幸福感的大幅上升是一个很有吸引力的解释。

在此，我对数据进行的非正式检验为过去几十年中美国男性和女性的年龄对幸福感效应的方向和大致特征提供了有力的证据。这些效应的大小是很难进行精确测量的，因为它们在截面数据中与世代效应相混淆，且在世代内趋势数据与时期效应相混淆。我们可以使用统计模型来获得准确的估计，但是估计将有很大的误差，而只有当它们与根据非正式的方法获得的大致模型是一致的时候，估计才是可信的。另外，根据精确估计获得的结果并不一定比利用有关效应的方向和大致模式的信息更加有用。

在对年龄和幸福感的全面研究中，对估计出来的效应提供理由和相关的证据自然是非常重要的。在此，我简单地指出，年龄效应对男性和女性幸福感效

应的不同模式的一个可能的原因就是，男性和女性在同一年龄所处的婚姻状态可能不同。这种相关可能部分归因于选择结婚或者不结婚的人群的特征，但这当然在一定程度上是结婚对幸福感的正效应的结果。在较年轻的时候，女性比男性更倾向于结婚，但是在年纪较长时，男性比女性更倾向于结婚。根据1972年至2002年的美国综合调查数据，男女结婚人数比例的差异在19岁到76岁之间与年龄有着极大的线性相关（$r=0.95$，以每一年的年龄水平作为分析单位）。男女表明自己“非常幸福”的人数比例的差异同样与年龄有着较强的线性相关（$r=0.792$），而男性与女性在结婚比例和表明自己“非常幸福”比例的差异之间的相关系数为0.823。同样是用每一年的年龄水平作为分析单位，且分析的样本年龄在19岁到76岁之间（为了使每个单元格中的样本数量大于100），对男女表明自己“非常幸福”比例的差异根据年龄进行回归的回归系数为0.253（$p<0.1$，双尾检验），但是在加入男女结婚人数比例的差异后，年龄的系数下降为0.033（下降了87%）。当控制年龄时，对男女幸福感的差异根据结婚比例的差异进行回归的回归系数为0.263（$p<0.01$，双尾检验）。

另一种解决这个问题的方法是，对男性和女性分别根据年龄对幸福感进行回归，之后加入婚姻状况这一协变量，比较它对回归系数产生的影响。当对整个1972年至2002年的数据进行回归时，男性和女性样本中年龄的系数分别为0.004和0.001，它们的差别为0.003（$p<0.001$，双尾检验）。在这个模型中加入了婚姻状况之后，男性样本中年龄的系数下降为0.001，但女性样本中年龄的系数则几乎没有变化（因为女性的年龄与婚姻状况的线性相关几乎为0）。

这些数据显示（尽管不能完全的证明），男性与女性在婚姻状况上年龄的差别在很大程度上导致了男性与女性的年龄效应对幸福感影响的不同。

第 3 章 | 用世代分析理解变革

以下的世代分析仅限于人口的出生率、死亡率以及老龄化速度没有大的变化的同代。在这样一代相对封闭的人群中，从时间点 A 到时间点 B，群体特征的变化主要通过 3 种形式：(1)通过从初生到成熟的新增人口；(2)通过老龄化和死亡减少的人口；(3)在时间点 A 与时间点 B，除去由老龄化以外的其他变化所引起的群体中个体的变化。前两者构成“世代演替”，而最后一种叫做“世代内的演变”。[3]

在之前的研究中(Glenn，1977)，我认为将人群按照上述 3 种变化模式细分，对于分解研究人口是非常有帮助的，并且建议发展相关的研究方法。至少已经有一个研究者格伦·菲尔鲍，响应了我的提议并且设计出了几种将人口按照上述变化分类的方法(Firebaugh，1989、1990、1992、1997)。

在这里，我不对菲尔鲍的方法进行具体解释，因为我现在相信，无论是它们还是任何近似的分解方法，对于理解改变都不是很有帮助(Rodgers，1990)。这种分解研究仅对不具年龄效应的因变量有意义，或者在极为罕见的情况下，即当一个人高度自信地认为根本没有什么年龄效应时才有意义。这种研究的结果就是其太依赖于它所覆盖的时间范围了，所以没有意义。

用一个案例便可以证明年龄效应对于理解人口变化而言，多么没有意义：我们假设在 25 岁到 64 岁的人中，每衰老 10 年便获得＋10 的因变量。在时间点 A 与时间点 B 之间，年龄分布不发生变化，则从时间点 A 到时间点 B 只有＋10 的时期效应，并且随着时间的推移，“世代演替”也没有任何作用。从另一方面来说，年龄效应对于年龄分布不变人群的人口变化不发生影响，而对“演变”却存在某种影响(除时期效应以外)。因此，在这个假想的例子中，演变对于解释人口变化已经绰

绰有余。如果菲尔鲍设计的或任何相近的方法是实用的，那么它就应该展示世代演替对于人口变化显著的负面影响，尽管事实上它确实没什么作用。

另外一个问题是，显然世代演替在一个相对短的时期里对人群的变化几乎不起作用，可是如果时间足够长，则任何分解研究都会把人口变迁归因于世代演替。设想一下，同样以 25 岁到 64 岁的人群为例。40 年之后，没有一个初始成员还会留在这个群里，自然而然，在这里没有任何人口变化可以归因于演变。而如果我们一定程度地细化时间，比如，在 25 年到 35 年之后，这将会对由世代演替产生的人口变化以及解释可见的人口变化产生显著的影响（尽管它不应该产生这样的效果）。

即便如此，如果能得到合适的数据，世代形式对于理解社会、文化以及政治上的变迁总是有用的，例如，理解在已经发生的变化的影响下，由几十年的世代演替引起的人口不断演变。如果像某些理论以及大量证据表明的那样，通常时期效应对年轻人比对年长者更具有影响力，那么足够大的时期效应将会随之造成显著的人群内差别。

表 3.1　美国不同时期不同年龄人群对性行为态度的回归分析

时　期	婚　前	婚　外
1972—1976	—0.022***	—0.010***
1977—1981	—0.022***	—0.007***
1982—1986	—0.015***	—0.004***
1987—1991	—0.019***	—0.004***
1992—1996	—0.016***	0.001
1997—2002	—0.014***	0.000

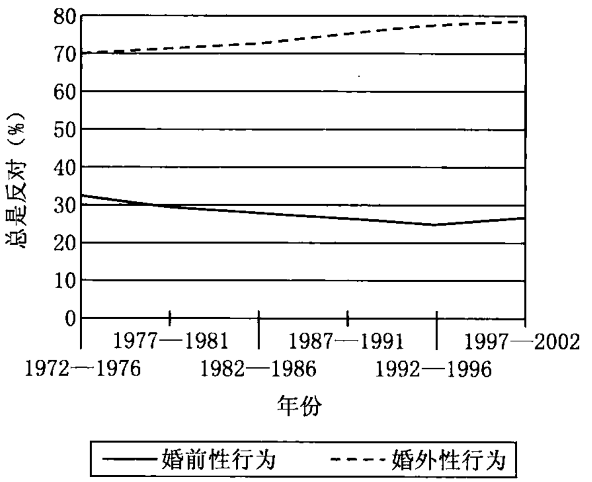
注：*** 表示 $p<0.001$（双尾检验）。

资料来源：由 1972 年至 2002 年美国综合社会调查的数据估算得出。负系数表明，年长者比年轻人更为保守。

20 世纪 60 年代至 70 年代早期美国“性革命”的影响似乎就说明了这样一种情况。在 20 世纪 70 年代中期，即这场运动的大发展时代，社会普查中关于性别差异的重复问题的答案出现了相当大的“年龄差异”：在行为模式上，年轻人比年长者更宽容自由（表 3.1），尤其在婚前性行为这一问题上。这些差异本来

已经能反映出年龄效应：年长者更倾向于在性行为方面对人们严格要求。不过，在下文所列的证据中，这看上去仅仅是由不同年龄的人对性革命的刺激有不同反应而引起的。

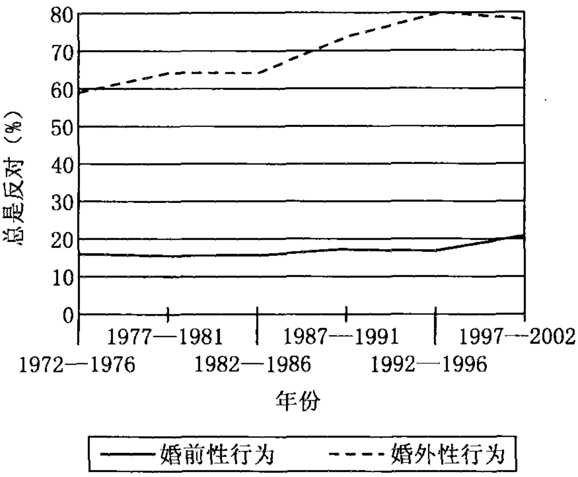
20 世纪 70 年代中期的这种人群内部差异有助于我们理解一个似是而非的说法，即从 20 世纪 70 年代中期开始到 20 世纪末，人们对于婚前性行为与婚外性行为呈现出截然相反的态度(图 3.1)，在所有的美国人中，人们对于婚前性行为的态度越来越宽容，而对待婚外性行为的态度则越来越严苛。这两种趋势在统计数据上都是非常显著的($p < 0.001$ 双尾检验)。这两种趋势看似互相矛盾，因为它们彼此相关($p = 0.301$)并且又以一种简单的形式和很多其他的变量相关。因此，如果时期效应对这两种发生于过去几十年内且截然相反的趋向产生影响，将会是非常奇怪的一件事。



资料来源：由 1972 年至 2002 年美国综合社会调查的数据估算得出。

图 3.1 美国不同时期对婚前及婚外性行为表示“总是反对”的 18 岁以上人群的百分比

时期效应不太可能以彼此反向的形式出现。正如笔者上文所述，时期效应的曲线通常体现在年轻人群以及 18—29 岁人群特点的曲线中。其实从 1972 年到 2002 年，人们无论是对婚前性行为还是对婚外性行为的态度都越来越严苛(图 3.2)，这种趋向在人们对婚外性行为以及婚前性行为的态度中非常显著，而在前者中体现得更为明显($p < 0.05$ 对于婚前性行为；$p = 0.001$ 对于婚外性行为)。



资料来源：由 1972 年至 2002 年美国综合社会调查的数据估算得出。

图 3.2　美国不同时期对婚前及婚外性行为表示“总是反对”的 18 岁至 34 岁人群的百分比

综上所述，几乎可以确定地说，由于世代演替，人们对于婚前性行为的态度将会变得宽松，并且这种世代演替的效果强大到可以忽略时期效应所带来的反作用力（更加保守）。与之相反，人们对婚外性行为的态度趋向于保守，并且时期效应的作用也足够强以至于可以排除世代演替在这方面的影响。这其中也许有一种趋向严格的年龄效应（年龄越大，态度越保守），也许没有，或者即使有，也会被时期效应或者世代演替的作用掩盖。

了解这些设定时间段内的人群内曲线总是有益的，1972 年至 2002 年这 30 年内的世代曲线在表 3.2 以及图 3.3、图 3.4 中都有体现。我使用根据人们态度宽严差异的二分法（总是反对相比于其他态度）进行回归分析。并且采用了顺序度量作为这张表上半部分的最小平方回归分析的区间。

表 3.2　美国不同年代的出生人群对性行为态度宽严对比的年度（1972—2002 年）分析

世代（出生年份）	婚前	婚外
一般最小二乘法[a]		
1925—1934[c]	−0.002	−0.008***
1935—1944[d]	0.002	−0.008***
1945—1954[e]	−0.012***	−0.011***
logistic 回归[b]		

续表

世代(出生年份)	婚前	婚外
1925—1934[c]	0.004	0.029***
1935—1944[d]	0.006	0.034***
1945—1954[e]	0.030***	0.035***

注:a. 负系数显示了越来越严苛的态度趋向。

b. 人们的回应被分为"总是反对"对"其他态度",正系数显示了越来越严苛的态度趋向。

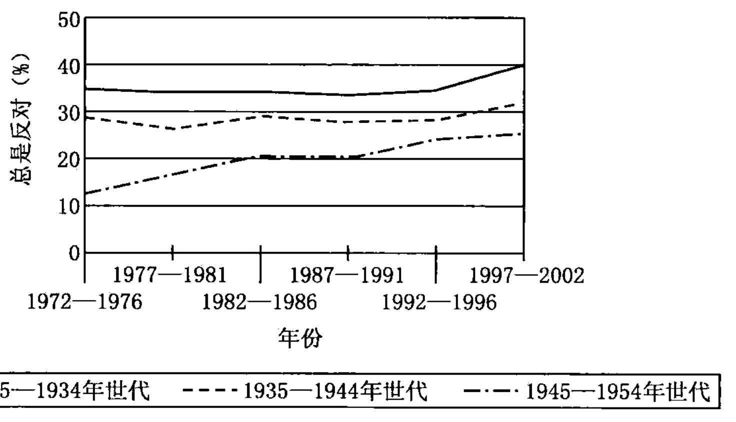
c. 本数据覆盖了年龄从 38—47 岁开始到 64—73 岁为止的世代。

d. 本数据覆盖了年龄从 28—37 岁开始到 54—63 岁为止的世代。

e. 本数据覆盖了年龄从 18—27 岁开始到 44—53 岁为止的世代。

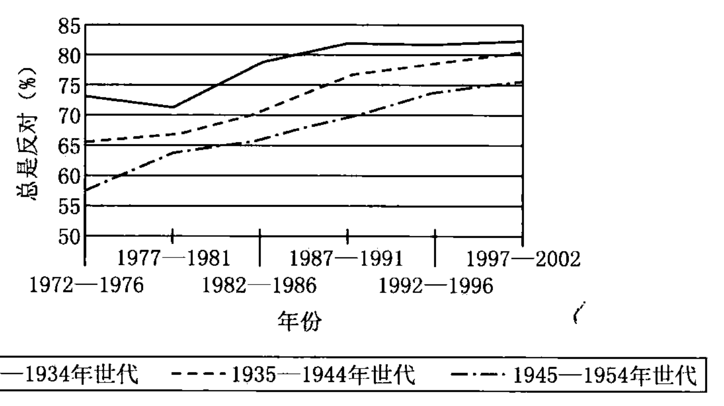
*** 表示 $p < 0.001$(双尾检验)。

资料来源:由 1972—2002 年美国综合社会调查的数据估算得出。



资料来源:由 1972 年至 2002 年美国综合社会调查的数据估算得出。

图 3.3 美国不同时期及不同年份出生的人群对婚前性行为表示"总是反对"的百分比



资料来源:由 1972 年至 2002 年美国综合社会调查的数据估算得出。

图 3.4 美国不同时期及不同年份出生的人群对婚外性行为表示"总是反对"的百分比

人群内婚前性行为态度曲线的意义并不是很清楚。只有最年轻的群体才在统计中显示出显著的趋向性，其他两个世代在 1992 年至 1996 年以及 1997 年至 2002 年的时间段里都趋向于严格。我和哈丁以及詹克斯(Glenn, Harding & Jencks, 2003)一样，已经通过简单的数据来说明人们从 30 岁开始就可能会产生一种年龄效应，从而在态度上趋向严格，但在 3 个世代中，这种简单的变化(当接近该世代所覆盖时间范围的终点时)却也很有可能是由时期效应造成的(表 3.3)。无论如何，人群内缺少宽容的态度趋向证明了群内缺少年轻人群的趋向，这也意味着在整个人群中，宽容态度趋向的缺失是由世代演替造成的，和时期效应没有关系。

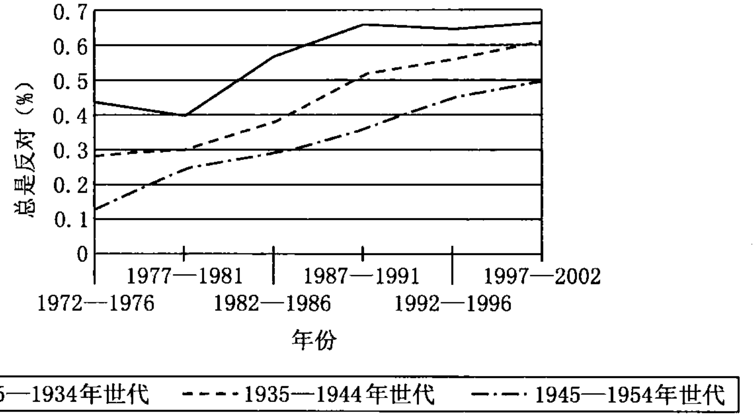
群内对婚外性行为的态度曲线的意义则更加清晰。3 个世代都显示出强烈而明显的越来越严苛的态度趋向(表 3.2 及图 3.4)，这样的变化很大程度上是由时期效应引起的，尽管年龄效应在这里不能被排除(随着年龄增长，态度趋向严苛)。在这里，趋向严格的时期效应必须足够强大，除了能够形成明显可见的趋向曲线，它还需要排除由世代演替所引起的在该分析早期所产生趋向宽松的反效应。在表 3.1 中，在“1972—1976”以及“1987—1991”的时间段内，年龄与采取严苛态度之间的关系显示了这样的反效应。

尽管如此，如表 3.1 所示，在 20 世纪 90 年代中期，年龄与采取严苛态度的对应关系却消失了。这个变化也许是由于青年人转而比年长者采取更加严苛的态度所造成的。同样，这个人群的新加入者(新成年者)所采取的态度与年长者更为相似也是原因之一。综上所述，这些发现其实与“对于变化的刺激，年轻人群更为活跃并容易作出应答”这个假设是一致的，但图 3.4 中的趋向曲线却没有显示出 3 个世代的单调收敛性。此外，由于天花板效应，最年长组比起年轻组，趋向变化要少一些，也就是说，由于该组成员的态度实际已经为相当高水平的严苛，这使得一些让人们态度趋向更加严苛的方法对他们已经不起作用。

当二分法被应用于一个连续变量时，即当把人们的宽松—严苛态度差异问题转化为“总是反对其他态度”的问题时，做一个关于“修正”上限及下限效应百分率的分数的对数转换是非常有益处的。这个转换包括将不平等百分比转换为发生比，并对比值进行对数转换。这个过程使得相对于接近 50 的值来说，接

近于0或100的“区别”占了更大的比重，并且它建立在连续变量的分布大体呈正态的基础上。当然，这种“修正”的正确性取决于假设的正确程度。

与图3.4中表现的百分比对应的对数转换结果展现在图3.5中。



资料来源：由1972年至2002年美国综合社会调查的数据估算得出。

图3.5　美国不同时期及不同年份出生的人群对婚外性行为表示“总是反对”的logit转换

在1972年至1976年间，该数据仍然体现出世代内差异的净减少，但它同时也体现出不同世代间的差异（1977年至1981年及1987年至1991年间）。因此，通常意义上来说，这样的数据过于“凌乱”以至于不能支持该假设。造成这种情况的原因是多种多样的，除了年龄效应被复杂化以外，还可能是因为年长者对于变化刺激的反应其实与年轻人相一致（只是不那么迅速）所造成的。

第 4 章 | 数据要求与数据可用性

数据要求

我们通常都在两个或更多时间点上来收集世代分析的数据，而且最有用的数据通常都覆盖几十年的时间跨度。很显然，世代分析不会按照教科书中的顺序来制定研究计划、收集数据和进行分析，因为它的数据形式和数据可用性都是由其他学者根据其他目的而收集的。换句话说，它们其实只是二手数据。反之，有用的数据通常都是从样本研究中直接获得的，尽管其中有些数据（比如人口普查资料）原本是针对所有人而非特定人群进行的。

世代分析的一个重要要求就是，所研究的人群要近似于“封闭”，也就是说，除了出生和死亡人口以外，人口流动性要较小，或者通过年龄进行分组，比如使用 25 岁至 64 岁的人群作为研究对象。用其他方法所导致的人口的进出可能使人口年龄特性变化出现假象，或者掩盖已经发生的变化。因此，地区、州（美国的州）、社区以及类似的地区不适合进行世代分析研究，那些已婚人士、雇员或者相互有来往的人们也不适合。尽管没有什么人是完全封闭的，但单一民族国家的成年人却异乎寻常地适于进行世代分析，除非这个国家经历过大规模的移民。根据性别、种族、民族甚至宗教进行分组的人口也同样足够封闭。当然，根据成年前的背景资料进行的人口分组在其成年后也是封闭的。

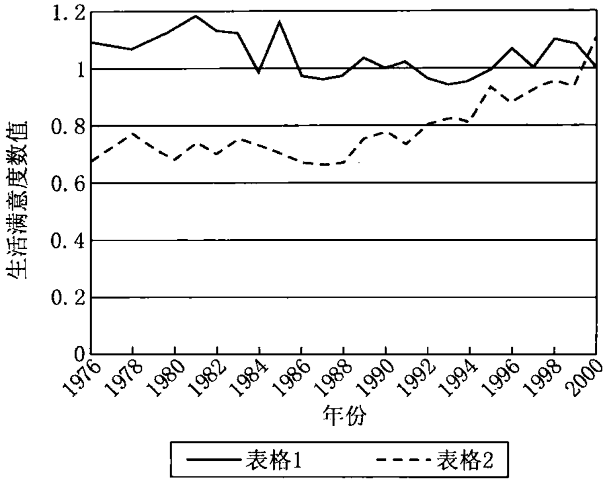
世代分析的第二个重要要求就是，在不同时间点所收集的数据必须具有可比性。比如调查问卷的数据，所有题干和可供回答的选项必须保持一致，除非有证据证明这些不一致不会导致反应类型的变化。问卷措词的微小不同可能导致反应类型大相径庭，所提供的选项顺序不同也会导致结果有很大的出入。

比如，在一个关于生活满意度的调查问卷中，调查对象被要求按照从“完全满意”到“完全不满意”的几个数值范围中进行选择。满意度排序的最先选项可能导致调查对象的选项发生变化。数值范围的中间选项是用言语表示还是用数字表示也可能会导致结果的不一致。

即使调查问卷的题干的措词和选项都一样，结果还是有可能有差别。由不同的问卷调查组织收集的数据可能因为样本设计、访谈人士的训练和监管、编码程序(机构效应)而不同，通过不同调查方式(面对面、电话、自行管理)所得的数据，严格意义上并不具可比性(模式效应)。更重要的是，早前被问到的问题，尤其是刚问过的问题，会极大地影响调查对象的答案(问卷语境效应)。比如，关于总体生活满意度的问题是否出现在关于某个领域生活满意度问题之前，所导致的选择很可能很不一样。

很明显，用可以比较的数据追踪世代或同年龄人的发展趋势作为相应水平的演替的世代是非常重要的。而不明显的是，即使可比较的数据也会给未发生的趋势制造假象或者掩盖真实趋势。这一点在图 4.1 中已经详细举例说明了。该图显示，在关于高中女生对生活满意度的调查问卷中，由密歇根大学社会研究所进行的“监控未来调查”中用两份不同的二次抽样样本所得到的选择结果存在组别趋势的差异。

在年度调研的表格 1 中，学生被要求使用包含 7 个数值的量表对其生活满意度进行评估，该数值范围将“完全满意”放在第一位，最后是“完全不满意”。数值范围的中间值是“中性”，数值范围的其他点则仅仅使用了数字而不是数字和文字并用。在询问了关于生活的 11 个具体方面的满意度之后，包括从“你的教育经历”到“你是如何打发闲暇时间的”，学生们被要求评价“生活整体满意程度”。在表格 2 中，调研人员先对二次抽样中独立抽出的学生们问了许多完全不相关的问题，然后询问他们：“最近你的生活整体满意度如何?”问题的选项先列出了“完全不满意”，然后才过渡到“完全满意”，并且这 7 个等级数值范围的中间值包含了数字和书面标签，比如“相当满意”或“还算满意”。在图 4.1 中，我给“完全满意”标上+3 分，“完全不满意”标－3 分，中间值标 0 分。每年每张表格的平均反应值都被绘制在数据中。



资料来源：由 1976 年至 2000 年"监控未来调查"问卷得出。以 7 分为生活满意度标准值，完全满意则加 3 分，完全不满意则减 3 分，表格 2 用文字标注的形式来标示所有 7 个问题。

图 4.1 美国高中女生年度生活满意度值

尽管表格 1 和表格 2 的问题在名义上是一样的，但不仅平均反应值不同，而且几十年来的发展趋势也朝相反的方向进行。当使用问卷总数进行分析时，表格 1 中的问题所得出的平均值的相关系数是 −0.458，是具有统计意义的（$p<0.05$，双尾检验），而表格 2 的平均值的相关系数是 +0.793（$p<0.01$，双尾检验）。很明显，这两个相关系数都不能准确指明美国高中高年级女学生生活满意度的最新发展趋势。

出现这两种相反趋势变动的原因尚不知晓，因为这两份关于生活满意度的调查问卷有 3 个地方不同。第一也许是最重要的一处，表格 1 中关于生活整体满意度评价的问题是在问了生活的 11 个相关方面满意度的问题之后提到的。第二，表格 1 的该选项是按照"完全满意"在先，过渡到"完全不满意"的。第三，表格 1 仅在局部范围内用数字来标示 7 个选项中的 4 个，而表格 2 是用文字形式标示所有 7 个选项。

这种对两个问题应答的不同趋势很有可能（虽然不确定）是由提问顺序效应引起的，因为对于表格 1 内一部分特定问题的回答显示出向下的趋势。也就是说，前述问题可能引起了人们对满意度下降的生活各方面的思考。在早期被

回答问题中的一种趋向可能会产生一种延续该趋向的表象或是假象，但如果在后续阶段受到提问顺序效果的影响，则有可能产生变化。

从这个例子中我们了解到，如果对因变量的测量很容易受到问题顺序效应的影响，则应该在世代分析中尽量避免这类测量。如果使用了这种方法，得出的分析结果应当标注明显的警示。

由于样本设计的变化，由同样的机构按照同样的形式（例如面对面调查）在不同时期收集的数据也许是不具有可比性的。正如我在下文中将指出的，在20世纪30年代、40年代以及50年代所进行的那些民意调查对于分析趋向以及世代分析而言，都是有用的资源，但不幸的是，早期民意调查所收集的那些数据与新近的数据相比，不具有可比性，即使是由同一个机构收集的数据。早期调查是以控制配额的方式来设计样本，并按照由年龄、性别、地区对不同人群进行配额。例如，一个调查可能需要10个男性受访者、10个女性受访者、6个35岁以下的受访者、8个35岁至39岁的受访者以及6个60岁及以上的受访者。考虑到这些限制，调研者就已经确定了寻找目标受访者的大体方向，这个致命的弱点会导致严重的偏误。到20世纪50年代后期，尽管全概率样本在美国人口普查局以及类似的学术机构中（如密歇根大学的调查研究中心）还没有被使用，但大多数民意调查都已经采用了更为精密的样本设计方法。最典型的例子是，一种类似于全概率样本的分阶段地区样本过程被用于街区层级的抽样，但是在选定受访者的最后阶段，配额选择仍然是重要的方法。这种形式的样本被称做“改良概率样本”或“配额概率样本”，与已经确定了受访者选择方向而且没什么选择自由度的配额控制样本有重大的区别。严格地说，用这样的方式取得的数据要么与配额数据样本不相容，要么与全概率样本不相容，但很显然，它们与后者更可能接近一致。

美国民调抽样方式最重要的改变出现在美国盖洛普民意测验机构，这个机构早年叫做“美国民意调查研究所”。盖洛普第一个伟大的成就是预测了1936年总统选举的结果，在此后的20年间，这个机构一直拥有敏锐的政治眼光并主要关注总统选举。盖洛普早期的样本自然也是为代表选民而不是同代人群设计的。在这样的样本里，那些选举权较低的选民明显代表名额不足，像妇女、黑

人以及没有受过高等教育的人。自20世纪50年代早期开始，盖洛普的样本开始改良，这个过程直至1958年。此次改良计划使选择其他应答的受访者的百分比上升8到10个百分点。

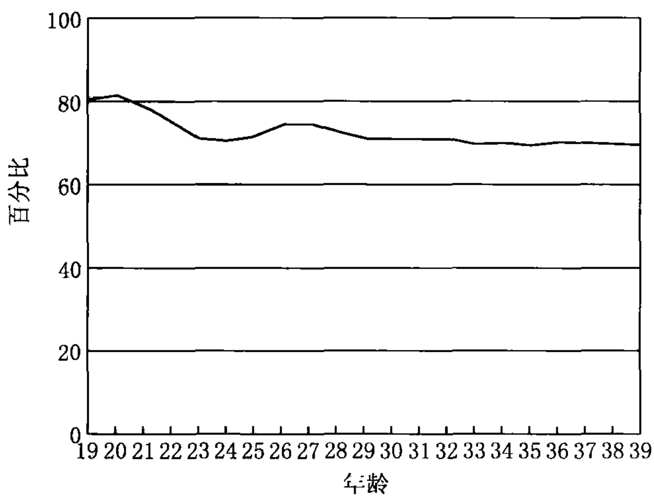
盖洛普所收集的数据曾经涵盖了大量趋向研究和世代分析可用的数据，我曾经建议采用多样化标准程序以使盖洛普早期和晚期收集的数据相匹配。这些数据偶尔会很有用，但随着不同时期、不同主题(社会学家们感兴趣的)的高水平学术调研所收集的数据越来越多，盖洛普的数据已经不像以前那么有吸引力了。因此，笔者在此不再讨论采取技术手段使盖洛普早期和晚期收集的数据相匹配的问题。大多数世代分析所用的数据也不存在这个问题。

还有一些问题不像盖洛普早—晚期数据匹配这么严重，它们是由主流民意调查机构将其调研方式从面对面改为电话采访而引起的。这些问题在20世纪70年代末到80年代初尤为典型，但它们并未严重到不足以影响对于同一机构所收集的早期数据和晚期数据的运用，但研究人员必须了解什么时候该用电话调研方式，并且要寻找那个时期对一些重复问题的应答的急剧转变。

在美国的大多数民意调查中，年轻人并没有像年长者那样充分被代表，由此引发的一个可比较性问题经常被人们忽视(包括我)。几乎所有的主流民意调查样本都只研究没有被制度化的人群，而“成群驻扎”并没有被列入研究对象，它包括监狱、医院、军营以及学校宿舍等地方，除医院以外，此类地方都有很多年轻人。这种由于上述群组年轻人代表名额不足所引起的偏倚，在某种程度上是可抵消的，因为从贫困的学生到中高阶层的学生都被排除在样本之外。但是，对某些变量来说，这种偏倚是不可忽略的。

研究世代分析的学者应当注意这种潜在的变化并注意这种存在与成年期早期及中晚期之间的变量的变化。不过，这个问题对大多数世代分析来说，似乎并不重要。老年人与年轻人世代是否可以进行比较，则可以通过观察随着年龄的增长，世代内背景特征是否相应变化来进行评估。因为有这些变量，任何群体内的潜在变化不是样本错误的结果，就一定是样本间没有可比性的结果。基于这种估计，运用第2章所提到的关于幸福的程序论证以及对同样年龄范围内(不是同样时间范围内)的个体进行以1年为单位的跟踪调研是很有用处的。

我运用这个程序，分析了几个未成年背景特征，没有发现这种人群内表现出性观念的变化（随着年龄增长）。在图 4.2 中，我们可以看到这样一个例子：自 1955 年至 1964 年，在美国出生的世代从 30 岁开始就被追踪调研，并且我们假设这个人群的个体中，只有一位接受过高中教育的母亲。在此背景特征下，表示性观念变化的百分比自个体接近 20 岁到 25 岁左右的阶段出现了一定程度的下降，这大概是因为大学生回归到了样本人口的范围中，并且在 30 多岁时依然保持该状态。这种已经显示出来的细微变化对于世代分析的结果不会产生很大影响。尽管如此，那些对于年轻人年龄效应没什么特别兴趣的学者还是倾向于用 25 岁而不是 18 岁作为它们分析的年龄起点，以降低由年龄相关样本的不可比性而引发的偏倚的概率。



资料来源：由 1972 年至 2002 年美国综合社会调查的数据估算得出。

图 4.2 美国 1955 年至 1964 年出生的世代中受访者母亲接受过高中教育的百分比（根据年龄，每 3 年计算一次均值）

数据可用性

本书中，我在例子中引用的数据大多来源于美国综合社会调查（在 1972 年至 2002 年间每年或每两年公布一次）。许多数据并没有覆盖该时间范围，并且

我并没有使用多次重复发问的问题，还有一些从1973年开始或是覆盖时间范围很短的问题也是如此。在这些问题中，有的问题在每个调查中都会出现，而有的则以一种螺旋式的方式呈现，例如在两个前后演替的调查中被提出，结果该问题在下一个社会调查中被保留，然后又出现在两个新的前后演替的调查中。近年来，社会普查数据每两年公布一次，其中的样本自然是早些年样本大小的两倍，并且这些数据涵盖了很多只有一部分受访者会被问到的问题。

美国综合社会调查中的一些更深入的问题涉及对社会生活各方面的满意程度、对当前形势的信心、思想意识取向、政党认同感、投票参与总统竞选的情况、工作满意度、工作态度、疏远感、志愿者组织成员资格、对孩子成就的期望、对堕胎的态度、词汇量、宗教偏好、宗教信仰、民族态度以及吸烟、饮酒状况等。和大多数面对面调查以及电话调查一样，美国综合社会调查的样本中有很多"家庭"代替了"个人"，因此，这样的样本必须从家庭中个体数量的角度被重新加权，以确定它能够代表被选定的世代（没有被重新评估的数据过高加权了只有一个成年人的家庭中成年者的数量，过低加权了多个成年者家庭中成年人的数量）。为了保证研究中显示的受访者与实际受访者数量一致，我们使用一个部分加权而不是原始的家庭中成年者数量。这个加权平均值由分割家庭中的成年者的数量（美国综合社会调查中得到的数据）得出，为1.94（1972年至2002年间）。这个权数估算可以在每一个调查中进行，但它通常只在由所有调查数据构成的汇总文件中才能进行。在1982年和1986年的调查中，黑人的过度抽样中出现了一个可变权数，但随着美国综合社会调查的范围越来越大，黑人的叠加采样变成相对非常小的一个部分，对此进行加权对结果几乎没有影响。纯粹的统计学理论研究者可以加权给每一个研究得出相等的权数，但这么做其实是拘泥于细节而没什么应用价值的，并且它对整个研究结果的影响可以忽略不计。

美国综合社会调查的数据对于很多研究都有其价值，从洛普公众意见研究中心的光驱到社会及政治研究校际财团的网站上都可以获得，带有硬盘截屏图的电码本还能在洛普中心买到。美国选举研究（ANES）比美国综合社会调查发出了更多的重复提出的问题，包括好几个系列，比如是否缺乏政党认同感等等，

这个问题从1948年第一次ANES进行以来，就一直包含在调查之内。其他几个"系列"从1952年开始，并且这些问题所覆盖的时间长度比美国综合社会调查中覆盖时间跨度最长的问题还要长。这些问题大多是从政治视角出发的参与性很强的问题，它们为政治理念提供支撑，甚至有很多党派偏见性的问题，社会学者、社会心理学者、经济学者，还有政治家对它们都很感兴趣。例如，ANES问了许多关于政策方面的问题，涉及医疗保障、民权行动、堕胎以及军事、经济等方面。

ANES有自己的网站，通过选择所需信息表格的引导，你可以在上面下载你需要的数据，并且加权信息也是给出的。在进行世代分析或者其他覆盖时间范围较久的社会分析以前，你必须好好研究这些复杂的信息。有些加权信息是从1994年才开始提供的，并且如果你要使用早年的数据，你就会发现这些不同时期的数据完全不具有可比性。

除美国之外，一些其他国家也在实践类似于美国综合社会调查和ANES的信息采集工程。我们可以从社会及政治研究校际财团的网站上找到其中的一些数据，但有一些只能在美国以外的档案馆或网站上找到。

一个真正能给世代分析带来巨大好处并且已经收集了很多有用信息的信息采集工程是"世界价值调查"，一个关于社会文化与政治变迁的世界性调查。第一次世界价值调查从1981年至1984年，其后还有1990年至1993年、1995年至1998年以及2000年至2001年3次，每次调查都至少有来自80余个国家的超过1000名的受访者参与。有20余个国家参加了所有4次调查并且提供了重复并具有代表性的数据，我们可以用这部分数据来进行世代分析。这项调查提出了一大批多种多样的问题，如环境问题、社会规范问题、对婚姻及家庭的态度、性观念问题以及政治态度等等。前三次世界价值调查的数据可以从社会及政治研究校际财团的网站上找到，并且随着本书的完成，第四次调查的数据将被提供给大众。关于这项调查的最新信息可以在世界价值调查网站上找到。

尽管有其局限性，但商业民调数据对于世代分析，仍是一种十分有用但却未被充分使用的数据，它所提供的数据周期很短而不像一般社会普查那样是1年至4年(在此期间有很多重复的学术研究)。在美国，很难想象有什么主题不

是以民意调查形式来进行的，并且有很多好的问题被重复发问了数年甚至几十年。民意调查主要由两个机构负责：康涅狄格大学的洛普公共意见研究中心以及北卡罗来纳大学教堂山分校中由奥德姆学院负责的路易斯·哈里斯数据中心。洛普中心是世界上历史最悠久也是最大的民意调查中心，它保有自 1935 年以来由民意调查机构所收集的绝大部分非专利数据。人们可以通过洛普中心的网站来搜寻数据，但数据仅对中心成员以及缴纳使用费者开放。另外一个主要的民意调查机构——路易斯·哈里斯中心——所收集的大部分数据，仅对奥德姆学院可用，这也使得后者保存了自 1958 年开始的超过 1200 个由哈里斯中心所做的民意调查的数据。人们同样可以在奥德姆学院的网站上搜寻数据，并且某些数据是免费的。其他各工业化国家至少拥有一家保存了丰富民意调查资料的机构。

不幸的是，年龄仅仅在某些早期民意调查的大目录下被编码，这使得它们对于世代分析的价值非常有限。即使要为世代分析寻找合适的民意调查数据，但在消耗大把的时间和进行数据研究之前，人们也应当看看年龄是不是编码的。

第5章 | 未来展望

世代分析的前景是光明的，因为有大量的可用数据能够支撑这种研究继续前行，并且，现在其实已经有大量的数据可以帮助我们理解年龄、社会以及文化的变化。尽管不会有一种绝对理想的方法被应用于“年龄—时期—世代”难题，但仍然会有方法论上的不断改良，从而推动世代分析研究的发展。

采用精密的统计方法而不是本文所涉及的简单方法会对世代分析产生积极的作用，尽管这些方法仍不能解决“识别”问题。我们应当承认，这些努力对于其他社会研究和世代分析来说，都有其意义，但显然专门为世代分析而设计的研究方法一定会更有效。

不幸的是，在短时期内，这些其他的研究方法对世代分析的意义是有限的，因为分析中既要用到过去的数据，又要用到新近的研究结果。世代分析依然会受到“所使用方法不够先进”这样的批评。尽管如此，世代分析自身的高敏感性对解决其方法论问题仍然会有所帮助。

目前为止，几乎所有的世代分析都会用面对面调查这种形式以获得数据，但几乎可以断定的是，这样的问卷调查以后会更依赖于其他方式以获得数据，例如通过电话调查来获得数据。这些以其他方式所获得的数据会引起一些问题，而笔者在专著中并未提及或仅仅是一笔带过。例如，与面对面调查相比，与年龄相关的样本的不可比性在电话调查中可能是不同的。如上文所述，不可比性问题是由数据使用引起的，在同样的分析中，用不同的管理问卷的方法能够获得不同的数据，并且总是会有诸如电话调查的数据不同期之类的问题。为了进行电话调查，应答机与来电者身份确认制度的广泛运用可能会引起新的偏

误，并且目前还没人知道这种对家庭电话进行调查所获得数据的依赖性增加会引起什么后果。

世代分析所面对的逻辑难题始终一样，然而它所涉及的技术手段却一定会改变，而且会在未来的几年里迅速改变。

注释

[1] 梅森和其他人没有指出在一个相对独立的可变年龄的范围内，对于一个线性变量，可能存在一个三变量解释。

[2] 人们可能会奇怪，为什么像群内成员父母离婚比率这样的数据也被用于个人层面的研究。这样做是基于一种假设的关联效应：父母没有离婚的儿童与青少年会受到父母离婚的同辈的影响，当然还有年龄关联效应以及时期关联效应。

[3] 年龄效应有时会被错误地列为可变因素之一。事实上，只有当群内年龄分布改变时，年龄效应才是相对可变的，这种情况在出生组（适合进行出生组分析）中随着队列演替而发生。

参考文献

Abramson, P. R. , & Engleghart, R. (1995). *Value change in global perspective*. Ann Arbor: University of Michigan Press.

Alwin, D. F. (1991). "Family of origin and cohort differences in verbal ability." *American Sociological Review*, *56*, 625—638.

Alwin, D. F. , Cohen, R. L. , & Newcomb, T. M. (1991). *Political attitudes over the life-span: The Bennington women after fifty years*. Madison: University of Wisconsin Press.

Bachman, J. G. , Johnston, L. D. , & O'Malley, P. M. (1976—2000). *Monitoring the future: Questionnaire responses from nation's high school seniors*. Ann Arbor: University of Michigan Institute for Social Research. Twenty-five volumes(Ordering of compilers' name varies.)

Blalock, H. M. , Jr. (1966). "The identification problem and theory building." *American Sociological Review*, *31*, 52—61.

Blalock, H. M. , Jr. (1967). "Status inconsistency, social mobility, status integration and structual effects." *American Sociological Review*, *32*, 790—801.

Converse, P. E. (1976). *The dynamics of party support: Cohort analyzing party identification*. Beverly Hills, CA: Sage.

Davis, J. S. , Smith, T. W. , & Marsden, P. V. (2002). *General social survey, 1972—2002*(machine readable data file). Chicago: National Opinion Research Center.

Firebaugh, G. (1989). "Methods for estimating cohort replacement effects." In C. C. Clogg(Ed.), *Sociological methodology 1989*(pp. 243—262). Oxford, UK: Basil Blackwell.

Firebaugh, G. (1990). "Replacement effects, cohort and otherwise: Response to Rodgers." In C. C. Clogg(Ed.), *Sociological methodology 1990*(pp. 439—446). Oxford, UK: Basil Blackwell.

Firebaugh, G. (1992). "Where does social change come from? Estimation the relative contributions of individual change and population turnover." *Population Research and Policy Review*, *11*, 1—20.

Firebaugh, G. (1997). *Analyzing repeated survey*(Sage University Paper series on Quantitative Application in the Social Science, 07—115). Thousand Oaks, CA: Sage.

Glenn, N. D. (1974). "Aging and conservatism." *Annals of the American Academy of Political and Social Science*, *175*, 176—186.

Glenn, N. D. (1975). "Trend studies with available survey data: Opportunities and pitfalls. In Social Science Research Council Center for Social Indicators." *Survey data for trend analysis: An index to repeated questions in U. S. national surveys held by the Roper Public Opinion Research Center*(pp. 6—48). Williamson, MA: Roper Center.

Glenn, N. D. (1976). "Cohort analysts' futile quest: Statistical attempts to separate age, period, and cohort effects." *American Sociological Review*, *41*, 900—904.

Glenn, N. D. (1977). *Cohort analysis*(Sage university Paper series on Quantitative Applications in the Social Sciences, 07—05). Beverly Hills, CA: Sage.

Glenn, N. D. (1980). "Values, attitudes, and beliefs." In O. G. Brim, Jr. , & J. Kagan(Eds.).

Constancy and change in human development (pp. 596—640). Cambridge, MA: Harvard University Press.

Glenn, N. D. (1989). "A caution about mechanical solutions to the identification problem in cohort analysis: Comment on Sasaki and Suzuki." *American Journal of Sociology*, *95*, 754—761.

Glenn, N. D. (1994). "Television watching, newspaper reading, and cohort differences in verbal ability." *Sociology of Education*, *67*, 216—230.

Glenn, N. D. (1998). "The course of marital success and failure in five American ten-year marriage cohorts." *Journal of Marriage and Family*, *60*, 569—576.

Glenn, N. D. (2003). "Distinguishing age, period, and cohort effects." In J. T. Mortimer & M. J. Shanahan(Eds.), *Handbook of the life course* (pp. 465—476). New York: Kluwer Academic/Plenum.

Harding, D. J., & Jencks, C. (2003). "Changing attitudes towards premarital sex: Cohort, period, and aging effects." *Public Opinion Quarterly*, *67*, 211—226.

Kiecolt, K. J., & Nathan, L. E. (1985). *Secondary analysis of survey data* (Sage University Paper series on Quantitative Applications in the Social Science, 07—053). Beverly Hills, CA: Sage.

Mason, K. O., Mason, W. M., Winsborgough, H. H., & Poole, K. W. (1973). "Some methodological issues in the cohort analysis of archival data." *American Sociological Review*, *38*, 242—258.

Nakamura, T. (1982). "A Bayesian cohort model standard cohort table analysis." *Proceedings of the Institutes of Statistical Mathematics*, *29*, 77—97(in Japanese).

Nakamura, T. (1986). "Bayesian cohort models for general cohort table analysis." *Annals of the Institute of Statistical Mathematics*, *38* (Part B), 353—370.

O'Brien, R. M. (1989). "Relative cohort size and age-specific crime rates: An age-period-relative-cohort-size model." *Criminology*, *27*, 57—78.

O'Brien, R. M. (2000). "Age-period-cohort-characteristic models." *Social Science Research*, *29*, 123—139.

O'Brien, R. M., Stockard, J., & Isaacson, L. (1990). "The enduring effects of cohort characteristics on age-specific homicide rates: 1960—1995." *American Journal of Sociological*, *104*, 1061—1095.

Palmore. E., & Jeffers, F. C. (1971). *Prediction of life span*. Lexington, MA: D. C. Heath.

Rodgers, W. (1990). "Interpreting the components of time trends." In C. C. Clogg(Ed.), *Sociological methodology 1990* (pp. 421—438). Oxford, UK: Basil Blackwell.

Sasaki, M., & Suzuki, T. (1987). "Changes in religious commitment in the United States, Holland, and Japan." *American Journal of Sociology*, *92*, 1055—1076.

Sasaki, M., & Suzuki, T. (1989). "A caution about the data to be used for cohort analysis: Reply to Glenn." *American Journal of Sociology*, *95*, 761—765.

U. S. Census Bureau. (1969). *Statistical abstract of the United States: 1969*, Washington, DC: U. S. Government Printing Office.

U. S. Census Bureau. (1979). *Statistical abstract of the United States: 1979*, Washington, DC:

U. S. Government Printing Office.

U. S. Census Bureau. (1990). *Statistical abstract of the United States: 1990*, Washington, DC: U. S. Government Printing Office.

U. S. Census Bureau. (1999). *Statistical abstract of the United States: 1999*, Washington, DC: U. S. Government Printing Office.

译名对照表

additive model	叠加模型
age cohort	年龄世代
age effect	年龄效应
Age-Period-Cohort model(APC)	年龄—时期—世代模型
Age-Period-Cohort-Characteristic	年龄—时期—世代—特征
birth cohort	出生世代
cohort	世代
cohort diagonals	世代对角线
cohort succession	世代演替
compositional effect	组成效应
control group	控制组
cross-sectional study	截面研究
identification problem	识别问题
intracohort change	世代内的演变
panel conditioning effect	条件习惯效应
panel study	固定样本跟踪研究
practice effect	练习效应
quasi experiment	准实验
question-order effects	问题顺序效应
quota control sampling	控制配额抽样
randomized experiment	随机实验
side information	附加信息
small birth cohort	小型生育世代
standard cohort table	标准世代表

纵贯研究

作者简介

斯科特·梅纳德(Scott Menard)

美国科罗拉多大学行为科学院副研究员。于1981年取得科罗拉多大学社会学博士学位。主要出版著作包括《人口研究》(*Perspective on Population*)、《问题青少年》(*Multiple Problem Youth*)、《童党》(*Juvenile Gangs*)、《应用logistic回归分析》(*Applied Logistic Regression Analysis*)。他现在的研究着重于从青少年过渡到成人的分析,包括青春期的经历和行为对成年期的影响。

译者简介

李俊秀

先后毕业于香港中文大学、纽约州立大学奥本尼分校。目前是香港科技大学应用社会经济研究中心项目经理。研究方向包括教育、健康和定量研究方法。

序

截面研究是分析某一时间点的数据,而纵贯研究则分析跨时间的数据。在典型的截面研究中,每个个案的变量都只在同一时间测量一次,如2000年美国总统选举前的民意调查。典型的纵贯研究是变量在不同的时期被反复测量。例如,同一批选民在1996年和2000年的总统选举前都被访问过,并回答同样的问题。第二种研究叫做“追踪样本研究”,一般来说,其研究成本比截面调查高很多。这样做值得吗?回答是值得,尤其是当衡量变化而且能广泛而确定地推断因果关系时。

玛丽·布朗(Mary Brown)教授是一名政治学家,假设她对经济投票理论很感兴趣,即选民个人经济状况的改变如何影响对总统政党的支持。我们假设在追踪样本研究中,时间1(1996年)和时间2(2000年)都问到受访者的财务状况和投票情况。时间1与时间2的变化代表态度和行为的真正改变。此外,由于时间顺序,它们能更有力地推断经济环境对投票行为的影响。至于因果关系的论据,追踪样本方法比截面研究强得多,因为截面研究无法记录真正的变化。

追踪样本方法只是纵贯研究中的一种,事实上,纵贯研究分为前瞻性和回顾性两种,梅纳德教授在这本书中都详尽地描述了不同的纵贯设计类型。除了追踪样本方法,还有总人口设计、重复截面和旋转追踪样本方法。总人口重复研究,如美国人口普查;重复截面研究是独立和反复的抽样,如美国国家选举研究;旋转重访法,即反复抽样,而且每隔几段时间会加入或删除子样本,如国家犯罪调查。不同的研究设计收集纵贯数据时,都出现一些特别的问题。例如追踪样本研究的个案流失可能很严重,难以推断变量在不同时段的变化。缺失数据的处理则更困难。本书会仔细地讨论这些问题。

纵贯研究的分析相当复杂,但梅纳德教授的解释非常简明。良好的分析策略部分取决于样本的形态,包括个案和时期的数量。少量个案和多时段具有时间序列特点,一般可用部分自回归调整程序。相反,多个案和少时段具有追踪样本特点,可用改变值或滞后内生变量来分析。由于有越来越多的纵贯数据可用,因此研究者不断地推出崭新的纵贯分析方法。本书浅显易懂,研究发展趋势、历史变迁、态度改变或动态过程的学者将从中受益良多。

迈克尔·S. 刘易斯-贝克(Michael S. Lewis-Beck)

第 1 章 | 概述

从国家层面上，收集纵贯数据的历史已超过 300 年，它始于新法兰西（加拿大）[①]的定期人口普查，并从 1665 年至 1754 年在魁北克省继续进行。虽然它不是第一次人口普查，但是代表第一次的定期人口普查，而不是单一、独立没有规则的人口普查，后者早在公元前 1491 年的以色列就已存在了（Thomlinson, 1976）。其后，1749 年的瑞典、1769 年的挪威和丹麦、1790 年的美国也开始进行定期的普查。美国是非常特别的，因为它的纵向普查数据是从建国第一个 10 年一直延续到现在。从个人层面而言，早在 1759 年，巴尔特斯和内塞尔罗德以及沃尔和威廉姆斯最早使用了纵贯数据集（主要是个案和履历资料分析）（Baltes & Nesselroade, 1979; Wall & Williams, 1970）。第一次世界大战后，美国就开始了不同的长期儿童发展研究项目。1970 年后，出现了各种各样的社会和行为科学的纵贯研究。纵贯研究的快速发展证明了，研究人员和主要研究经费机构都认同它的重要性。每当问起纵贯研究的重要性，通常会着重于其研究设计和分析等质量方面。

对许多人来说，纵贯研究已被吹捧成灵丹妙药，它可用于建立时间顺序、测量变化大小并提供因果解释的依据。虽然纵贯研究的确有它的优势，但其成本昂贵并且存在其他困难。纵贯研究并不是必要的，即使是用来测试因果关系（Blalock, 1962; Davis, 1985），特别是在预先已经知道变量的时间次序（例如生物或遗传特征，如性别、种族、年龄）的情况下。纵贯数据绝不能解决拙劣的研究设计和数据分析的问题。

---

① 指法国人在加拿大建立殖民地的新法兰西时期。——译者注

在本书中，我用“纵贯”一词并不单指一个方法，而是一系列类似的方法(Zazzo，1967，引自 Wall & Williams，1970)。最好的学习方法就是将其与截面研究进行比较。纯截面研究中的每个人、研究对象、国家或个案只测量一次，每道题、概念或变量的测量都在单一时间或时期进行，每个个案的每个变量的测量只显现于某段足够短的时间内(所有个案和变量最好同时存在)。这种测量数据称为“同时发生”，也就是说，所有变量和所有个案都在同一时间发生。根据具体的研究，时间单位可以是秒、天、月、年或其他。社会行为科学研究中的“时间”与实验研究的“分钟”或跨国研究的“年代”不同。

纵贯研究的界定必须由研究数据和分析方法来判断。纵贯研究须满足以下条件：(1)每道问题或变量的数据要在两个或以上的不同时间收集；(2)不同时间的个案或对象要相同或至少是可比较的；(3)分析包括比较不同时间的数据。最低限度是，任何真正的纵向设计都可以测量从一个时期到另一个时期之间的变化或差异。根据这个定义，有几类研究可以被视为纵贯研究。其中一种是，在两个或两个以上不同的时期收集数据，每次都是同样的个案和变量，这就是前瞻式追踪样本设计。另外，收集同一时间和几个不同时期的数据，通常包括最后一期的数据，这是回顾追踪样本设计。回顾纵贯设计与前瞻纵贯设计在各方面基本上是相同，除了数据收集的次数和访问者的回忆时间长度。两个追踪样本设计在不同时间的个案和变量都要保持不变。第三种是数据在不同时段被重复收集几次，同变量但不同个案，这个方法是重复截面设计，每段时间的数据可被视为一个单独的截面数据，不过，因为个案可比较(例如使用概率抽样在同一人口样本抽样)，所以我们可以比较不同时期的变化。第 3 章将详细地介绍不同类型的纵贯设计。

巴尔特斯和内塞尔罗德以及沃尔和威廉姆斯提过狭义的纵贯研究，除了前瞻追踪样本设计外，他们排除了所有其他方法，但他们承认在定义纵贯研究上还没有达成共识(Baltes & Nesselroade，1979：4；Wall & Williams，1970：14)。巴尔特斯和内塞尔罗德认为应根据具体的研究情况去定义。心理学的发展研究用纵贯追踪样本设计可能较为适合，但其他学科又似乎过于严格。使用广义纵贯研究的原因，首先是对纵贯研究的定义缺乏共识，第二是考虑所有在

不同时期收集数据的方法的特点和实用性。

本书的第 2 章主要讨论纵贯研究的目的和历史发展变化的困难。第 3 章介绍并讨论收集纵贯数据的基本设计。第 4 章讨论可能影响纵贯数据质量的问题。第 5 章简略地介绍纵贯分析的方法。

第 2 章 | 纵贯研究的目的

纵贯研究有两个主要目的，第一是描述变化的模式，第二是建立因果关系的方向(正或负，从 Y 到 X 或从 X 到 Y)和幅度(零强度的关系表示没有因果关系)。测量变化通常涉及时间或年龄。时间是研究个案或对象的外部衡量(例如 2000 年 8 月 28 日下午)，年龄是研究个案或对象的内部衡量(例如从出生开始算起，经历 38 年 7 个月 26 天 8 小时和 27 分钟)。从某种意义上来说，年龄代表人类的生物时间。选择时间或年龄作为连续数可能是重要的，不过，为了某些目的，同时包括两者的分析可能很有用。区分与年龄相关的差距也是重要的，年龄数据是截面式(如，1990 年时 40 岁与 50 岁的差异)和纵贯式(如，同一个人在 1990 年是 40 岁，在 2000 年是 50 岁)。当测量年龄是截面式时，指的是 40 岁与 50 岁的变量差异，可解释为在某特定时间内，出生组或年龄组之间的差异。若是纵贯式，可解释为世代或年龄组随着时间发展而产生的差异。

年龄、时期和世代效应

年龄与时期在概念上是连续的，在度量和解释的变化上，可能会出现严重的问题。为了了解这些问题，在讨论测量的历史和发展变化之前，我们需要先讨论年龄、时期、世代效应之间的区别以及年龄、时期和世代作为变量和分析单位的不同概念的区别。

格伦定义了世代的人口结构(Glenn, 1979)，世代是指那些以地理或以其他方式划定的人，在某特定时期经历了同样的重大生活事件。莱德也提出过类似的定义(Ryder, 1965)。格伦和莱德都注意到，虽然世代经常用来指称出生的世

代(那些出生在某一年或时期的人),但是我们可以将其定义为某年结婚或离婚、退休、第一份工作、入学、大学或研究院毕业、发生事件的年份。格雷茨使用“事件世代”来描述出生世代之外的组群(Graetz, 1987)。

假设我们要研究人们的政治态度是否随着年龄增长而变得保守。方法之一是在某一年调查不同年龄的人,然后比较年龄较大与较小的受访者。如果老年人的政治态度比年轻人的保守,我们可以总结随着年龄增加,政治态度会变得更保守。但另一个可能的解释是,也许这些老年人年轻的时候就已经很保守,或许现在的年轻人以后仍然保持较不保守。换言之,截面研究结果的差距可能不是因为年龄因素,而是受到不同出生世代的影响,即不同的生活经验和年龄可能对人的态度有长远的影响。

现在假设不做截面研究,我们可以选择一个出生群组,从中选出部分人,每5年至10年进行一次访问,直到他们过世。研究结束后,如果我们发现受访者越老越保守,那么,我们可总结说,随着年龄的增加,政治态度会变得更保守。但是,仍有一个可能的解释,即在任何一年中,各年龄组都没有真正的差别,但每个人,无论是青年人还是老年人,都随着时间变得越来越保守,这跟年龄无关。这可能是历史、特定年代或时期的影响,而不是年龄。换句话说,无论年龄大小,当代事件可能对政治保守主义有实时的影响。一个时期的差异不会出现在截面的研究中,因为它只有一个时期。一个世代差异的问题也不可能在截面研究中出现,因为它也只有一次。

由于截面研究或单一纵贯研究都不能消除世代成员和周期效应的其他解释,所以比较合乎逻辑的方法就是把两者结合成多年多世代设计。然后,我们可以控制两者以考察年龄对政治保守主义的影响。问题是,当我们控制任何两个变量时,假设年龄、时期和世代成员的影响全是线性的,这是因为年龄、时期、世代的成员都是线性相关的,其中两个变量组成另一个线性函数。在数学上,

$$\text{世代(出生年份)} = \text{时期(年份)} - \text{年龄(自出生至今)}$$

我们测试政治保守随年龄而增加的假设仍受到干扰,例如把年龄、时期和世代都加入回归方程会导致完全共线性。任何年龄的影响,在没有控制时期和

世代的情况下，都可能是由于世代与趋势共同或任何一方的影响。格伦对于这个问题与世代分析的其他问题（抽样、样本死亡率等）都进行过详细的讨论（Glenn，1977）。

线性依赖与世代概念

尝试用不同方法去解决年龄、时期和世代的线性依赖包括，使用虚拟变量回归分析，同时限制模型中的参数的某些假设（Mason et al.，1973），或者重组和去掉 3 种效果中的一个或更多（Palmore，1978）。

这些方法，尤其是虚拟变量回归模型，是相当具有争议性的（Baltes et al.，1979；Glenn，1976、1977；Knoke & Hout，1976；Mason et al.，1976；Rodgers，1982a、1982b；Smith et al.，1982），包括假设的可行性和处理年龄、时期和世代的线性相关的后果。虽然虚拟变量回归模型中的限制可以消除完全共线性这个问题，但是模型中虚拟变量仍是高度共线性的，而且不同的限制可能产生非常不同的结论。此外，这些方法都不能解决线性相关的问题，因为梅森等人的模型假定，不是所有的效应都是线性的（Mason et al.，1973），帕穆尔（Palmore）的方法是预先消除一个效应。

请注意，线性相关的问题适用于出生世代，但对其他类别的世代不一定适用。在某种程度上，某个事件不会与年龄或时期相关，因为事件世代基于该事件，所以它同样不与年龄或时期线性相关。在某些情况下，线性相关可能预先已消除了。第二也是最根本的一点是，世代是把个人集合起来作为分析的单位。正因为这样，莱德做了一些世代影响的研究（Carlson，1979；Lloyd et al.，1987；Wetzel et al.，1987）。巴尔特斯等人讨论了世代的 3 个可能概念（错误或干扰；归纳维度；理论和过程变量）（Baltes et al.，1979），其中，归纳维度将世代作为分析单位多于理论变量。有些世代研究也将世代作为解释变量，这就认同了世代是一个分析单位（Wright & Maxim，1987）。

世代是将个人（个案）集合起来，分析方法也与其他研究（个人、城市、国家）一样。社会科学研究中的世代，像其他集合个案一样，都具有可量度的特性，如一些固有的累积性质（大小、性别比例、族群组合）等、总和（如被捕的总人数）或

平均数(收入中位数)。然而,我们不会这样量度年龄或时期,但我们可以衡量个案在特定年龄或时期中的整体特征。年龄和时期是时间聚集的总数,大多作为变量分析单位。在社会研究分析中,它们可能用来分开个案以进行分析,但本身通常不作为分析单位。

年龄、时期、出生世代的提问分别是:"你多大了?""今年是哪一年?""你出生在哪一年?"

"你多大了?"这个问题的答案也许可以解释某些行为模式——尿裤子是婴儿期最常见的,叛逆行为在青春期是最严重的,退休在65岁后是最普遍的。年龄是行为发展的一个解释。"今年是哪一年?"的答案也可能有助于解释某些行为模式——在美国,1960年后滥用药物比以前更普遍,1960年前种族歧视较常见。时期提供了一个解释,至少也是可能的解释(Hobcraft et al., 1982),其本质上就是历史,这些历史事件可能有助于解释某种特定的行为。"你出生在哪一年?"有两种回答方法。第一种叫做"东方占星术",即出生年份有某些特征。例如,根据东方占星术,出生在野马年(每60年一个循环)的妇女有杀夫的倾向。对于大多数社会科学家而言,这显然不是一个解释杀人的好方法,然而,无论是实际生育情况还是日本人口数据局的记录(但未必完全正确),都有同样的效应(Population Reference Bureau, 1989)。或许,哪年出生(成长)和某特定事件(历史)可能会形成某些行为。换言之,出生世代的效应用出生年份来衡量,可以视为年龄与时期的相互作用。

世代效应:重新定义和更换

研究世代影响的方法之一,就是把它作为年龄和时期的相互影响。另一种方法是假设不是世代成员,而是与出生世代相关的一些特征或特性而形成明显的世代效应。但问题是要确定该世代的适当特征或特点,这属于理论而非方法上的问题。

可能的解决方案是以衡量世代的方法取代出生年份,而用该出生世代的出生人数或世代的大小。莱德指出:"相对于邻近的世代,每一个世代的人口数量受到该世代所生存的环境特征的持续影响和制约。"(Ryder, 1965:845)梅森等

人指出，年龄、时期、世代是不可测量的变量指标，他们说："如果世代大小是造成（不同世代间）实质性差别的原因，并且可以通过测量得到该大小的话，那么就可以将世代大小作为一种更好的研究变量，而没有必要将世代（本身）纳入模型中。"（Mason et al.，1976：905）此外，用世代大小可消除虚拟变量回归模型的估计问题，因此，分析结果就不会太粗浅。莱德指出，世代大小是分辨各世代的特点之一，但是，自1968年出版了伊斯特林有关世代大小对劳工市场影响的研究，以及发表于1980年关于世代大小对社会问题，包括失业、离婚、罪案的研究后，世代大小已成了研究中一个很重要的角色（Easterlin，1987）。

当世代大小或其他世代特质（或涉及年龄与时期非线性相关）在概念或理论上适用于研究年龄、时期和世代效应时，我们有时会用出生年份来衡量世代。这样做至少可能减少了无法辨认年龄、时期和世代之间不同概念的情况所带来的部分问题。虽然世代成员从纯粹方法论的角度来看，可当做一个解释变量，但是年龄和时期更适合做解释变量，尤其是年龄（Hobcraft et al.，1982）。理想的情况是，可以删除时期和世代，然后换上可在因果分析中当指标的变量。

在研究发展或历史转变的分析中，应用多年多世代设计，加上适当地操控年龄、时期和世代效应（如应用世代特征或非线性年龄—时期交互作用），那么我们就可以测量年龄、纯时期和世代成员对政治态度的效应。其他解决年龄、时期和世代效应的混淆问题是理论上或方法上的问题，都需要在年龄效应出现之前找到。如政治保守主义的例子，理论上可能合理地删除影响政治态度的世代成员，或假设（基于理论架构）通过世代大小（或其他特质）去分析世代的成员。这样做才有可能估计年龄（成长）和时期（历史）对政治态度的影响。请注意，如果没有纵贯数据，发展、历史和世代成员的效应是不可能清楚地分辨出来的。

时期效应：随着时间而改变

只要有人想把年龄、时期和世代效应分开，就可能要检视它们随着时间而产生的变化。这需要不忽略世代效应，或用世代特征（如世代大小）代替。另外，如果我们只关心时间（历史的改变）而不理会年龄的变化（发展的改变），我

们就要否认年龄是完全无关的因素，或视年龄为解释变量，或控制年龄来进行特定年龄比较。

纵贯研究中的一个关注点就是简单地描述各变量随时间而变化的情况。在个人研究分析中，这可能包括宗教信仰、政治保守主义、饮酒情况。这些变化更常被视为个人发展变化多于历史趋势。在集合数据层面，我们可以使用犯案或受害者的比例、劳动生产量、人均国民生产总值、学业能力倾向测试(SAT)分数、婴儿死亡率、社会指标[1](测量基本社会需要上升或下降或达到合适的社会目标的情况)。在这个层面，可能需要控制年龄效应。婴儿死亡率已经表明特定年龄(虽然母亲的年龄会影响 1 岁前婴儿死亡的可能性)，SAT 的考生主要介乎 16 岁至 18 岁之间。因此，年龄已经完全或绝大部分被控制了。犯罪和受害的比例相当容易受到人口年龄分布的影响(Chilton & Spielberger, 1971; Skogan, 1976)。如果要评估变量的历史趋势，控制年龄是适当的。人均国民生产总值很容易受到年龄相关比例的影响(不足 15 岁、65 岁以上与 16 岁至 64 岁的比率)。同样，劳动生产量与劳工的年龄之间有相当高的相关性，但在收集数据后的一段时间内，这些变化会缓慢地变化或基本上保持一个常数，因此，可以忽略时期趋势中的年龄或年龄分布的变化。

一个最安全的研究趋势的方法就是使用特定年龄比较。在一个特定年龄比较中，只能将某年某个年龄的个案与随后几年的同龄人相比。年龄代表某个年龄(如 15 岁)或年龄组(如超过 65 岁)，以及分开比较所有可能的年龄或年龄组。例如，高德和他的同事在一个重复截面设计的研究中，检测了 13 岁至 16 岁青少年自我报告的犯罪率(Gold & Reimer, 1975; Williams & Gold, 1972)，结果发现，很少数据显示了 1967 年至 1972 年间的变化。

梅纳德的 1976 年至 1980 年国家概率抽样 15 岁至 17 岁的青少年的研究得到了类似的结果(Menard, 1987b)。科维和梅纳德探讨了 65 岁以上的受害和被捕趋势，发现在这一年龄组中，逮捕率普遍提高而受害率普遍下降(Covey & Menard, 1987、1988)。在上述例子中，控制了年龄后，有没有趋势的变化会相对明确些。如果没有控制年龄，即使用于整个人口而非一个样本，也可能难以确定变化本身是历史性的还是发展性的。奇尔顿和施皮尔贝格尔研究官方犯

罪率的变化，结果发现随着时间，大部分明显的变化（表面上可以改变人的行为）是由于年龄结构的变化，或者更具体地说，是人口中青少年的百分比（Chilton & Spielberger，1971）。个别的研究会有所不同，但一般情况下，明显的趋势变化可能是由于年龄的变化（个人层次）或年龄结构的调整（集体水平）。

变量关系的历史趋势

纵贯的另一个关注点是对变化的检验，这种关注不在于随着时间推移，变量数值或水平的变化，而是变量之间关系的变化。例如，死亡率在过去两个世纪以来一直下降。这说明公共健康措施（卫生、安全饮用水、杀菌等等）对早期死亡率的下降起了主要作用，药物发挥的作用有限，但后期阶段下降的主要原因是医学进步（疫苗、抗生素）多于公共健康措施（McKeown，1976；McKeown & Record，1962；McNeill，1976）。豪特等人研究从1944年至1992年，美国社会阶层（从专业到低技术蓝领）和总统选举投票行为之间的关系，发现不同阶层有不同的模式，最高层（专业）随着时间的推移，转而支持共和党候选人，最低的3个社会经济阶层（特别是非专业自雇人士和熟练工人以及技术人员）投票给民主党的趋势也随时间下降了（Hout et al.，1999）。

要探讨关系趋势变化的强度或方式，有一个重要的问题就是新数据能重复以前的研究结论。埃利奥特等人使用连续几年的数据去检验一个犯罪行为的理论模型，变量包括犯罪、滥用药物、精神健康（Elliott et al.，1989）。初期试验结果显示，该模型能很好地解释犯罪与滥用药物的关系，但不能很好地解释精神健康的问题。用同样的样本，一年后再衡量，埃利奥特等人成功地复制了第一次的结果。结果能够从一个时期复制到下一个时期（这是结果可信性）比单一没有复制的研究，为模型建构提供了更多的支持。

复制并不总是成功的，关系的强度或形式随时间而变化的证据可能表示关系有实际的变化或方法上出现问题，如不稳定或不可信的测量，或设定了错误的因果模型。梅纳德测试了一个生育率模型，这是85个发展中国家1970年至1980年期间的模型（Menard，1987a）。两个时期的整体形式都非常相似，但家庭计划生育方案之间的关系在1970年至1980年这段时间内有些变化，通常往

较弱关系的方向变化。除此之外，两个时期的模型结果几乎相同。正如梅纳德所指出的，这些变化可能反映了测量家庭计划生育的方案有所不同，但高的关系强度似乎暗示了这更可能是关系的真正变化而不是不可靠的测量。如果没有复制研究，无论测量的不稳定或不可靠性还是整体模型不同时期的一致性，都不可记录。

年龄效应:生命周期和发展变化

巴尔特斯及内塞尔罗德列出纵贯(更具体地说，对他们来说是前瞻式追踪样本研究)的5个目标或依据:(1)直接鉴定个体内的变化，即个人在某段时间是否有变化;(2)直接鉴定个体间变化的相似或个体内的差异，即个人改变的方式是否相同;(3)分析行为改变的相互关系，即变化是否互相相关;(4)分析个体内变化的原因或决定因素，即为什么个人在某段时间会有变化;(5)分析个体间变化相似或差异的原因或决定因素，即为什么不同的人在某时期有不同的变化(Baltes & Nessel-roade, 1979)。这些目标全部关注发展变化的形式，特别在个人层面上，虽然它们很容易伸延至集体层面(组群、组织、城市、国家)。在个人层面，个体内变化可能包括思想(政治越来越保守)、经历(就业、转工、退休)或者发生在他们身上的事(被逮捕或被抢劫)。在一项个体内变化的研究中，年龄是与年龄相关的生理变化和受社会影响的指标(Hobcraft et al., 1982)，要测量这些变化可能比较难。

基于某些目的，用截面数据简单地去推断个体内变化也是合理的。例如，从被捕率和生育年龄层的截面数据，我们可以合理地推断，7岁之前不可能被捕或生小孩，但在青春期和青年时，这种概率会一直增加，直到65岁后就又会大大减少。这种与年龄有关的差异不太可能是时期影响或世代特征。另一方面，基于截面数据而推断，人越老就越保守并接受越少的教育并不是很准确。如前所述，在特定时期，政治态度的年龄差异可能反映了态度随着年龄而变，或是整个生命周期世代差异保持常数。如果老一代人的教育比年轻一代的少，这不是因为他们“未受教育”，一个更合理的解释是，教育程度随着时间(时期效应)而增加，造成连续世代的平均教育水平有差异。

使用截面数据去研究年龄和行为之间的关系相当于建造一个综合世代，这是人口死亡率和生育率研究中常见的方法。例如，平均寿命和时期总生育率不是基于截面的死亡率和生育率，而是由保险公司或其他来推算随着年龄增长，有什么事件可能发生在个人或世代身上。如施赖奥克和西格尔提出，使用综合世代测量取决于在何种程度上，它们反映了世代的实际经验（可用对世代的纵贯研究进行评测）(Shryock & Siegel, 1976:324)。在某些情况下，截面和纵贯数据可能出现非常不同的发展模式结论。举例来说，格林伯格使用官方犯罪报告(Greenberg, 1985)，梅纳德和埃利奥特采用自我报告的犯罪研究，同时发现截面和纵贯数据对研究年龄和违法行为之间的关系可能导致不同的结论(Menard & Elliott, 1990a)。纵贯和截面结果的差异可能是由于世代大小的影响(Elliott et al., 1989:107—109; Menard & Elliott, 1990b)。

如果我们想研究行为的"职业"模式，就更迫切地需要纵贯数据。最明显的就是劳动力市场的研究，从最初入职、升职、转工、失业，最后到退休或死亡。类似这种研究包括教育程度以及职业地位和收入(Blau & Duncan, 1966)。其他"职业"角度的应用包括婚姻的历史(Becker et al., 1977)、教育程度、学习的过程(Heyns, 1978)和职业性犯罪(Blumstein et al., 1986)。这些研究的共同点都是关注行为的职业模式，从入职、继续和离开的行为，以及行为的转变或不连续性（失业、转新工、离婚、再婚、辍学和再入学、暂停和恢复犯罪行为）的相关性和潜在原因。只有纵贯数据，更具体地说，追踪样本数据，才可以回答很多关于职业发展模式的问题。

生命历程的研究(Giele & Elder, 1998)类似个人职业研究，但扩展了职业模式，明确地从更广泛的历史和社会背景角度去了解个人的变化。综合生命历程研究的角度如下：(1)时间性（历史）和地点（社会和文化）；(2)生命关系：从人际关系和社会制度的层面去看人与人的结合；(3)人类智能：个人目标设定和达到目标的能力和倾向；(4)掌握时机：决定何时采取行动或制定策略，不仅基于内在目标，还有外在的事件或条件。相反的观点认为，生命有固定的阶段，从生命历程的角度来看，个人的转变是由于个人目标（人类智能）和外在影响（掌握生命中的时机）的不同。生命历程着重现象研究，这只能从长期纵贯研究（事件

史或轨迹)中分析出来,不同个人的时机、期限或变动率都不一样。

纵贯数据在实验研究和评估研究中也很重要。大多数实验设计和准实验设计本质上是纵贯的,衡量实验性治疗或干预前(前测)后(后测)的变化(Campbell & Stanley, 1963)以确定变化是否由于干预的影响或实验组和控制组早已存在的差异。在实验设计中,即使没有前测,但假设研究者把参加者随机分配到不同治疗组,就能够令各组不会出现任何重要变量的差别或不相等的偏离,这是由于已知的统计分布。因此,即便后测的实验设计也包括一个关键的纵贯假设,即假设实验组和控制组于前测时没有或仅有很少差异,但前后测的差异只代表两组之差的差别。同样,评估研究通常会收集前测或基线数据(Rossi et al., 1999)。缺乏前测或基线数据的影响会导致无法确定治疗,或干预前后的变化完全基于治疗或干预的效应,或是两组早已存在的不同。

变量关系的发展趋势

除关注从一个时期到另一个时期的关系强度或模式变化,我们也想检验从一个年龄到另一个年龄的关系强度或模式变化。在这里,是基于截面数据的比较(世代与世代间)还是纵贯数据的比较(世代内)取决于我们是否关心能在截面数据上体现的发展变化。如果使用纵贯数据,则必须考虑任何变化是否由年龄、时期或世代效应所造成。

在美国新泽西州一个针对341位少年的研究中,拉格朗日和怀特发现,对年龄大(18岁)和年龄小(12岁)的青少年男孩,只有一个变量——与违法朋友的交往程度——对犯罪行为有重大的影响(LaGrange & White, 1985)。然而,对于15岁的男孩,家庭和学校的变量也影响到他们的犯罪行为,影响力有时比违法朋友更大。尽管样本数量很少,特定年龄的子样本甚至更少,但他们的研究提出的重点是,多元因果分析结果可能有所不同,至少在关系的强度方面,这取决于样本的受访者的年龄。由于数据是截面(同一年份的不同年龄组),就不可能排除另一种可能的解释:该年龄差异可能不是由特定年龄,而且是特定的世代造成。彻底的解决方案需要其他数据,最好是利用纵贯数据来重复验证该结论。

使用全国青年调查的数据,即1976年11岁至17岁和1986年21岁至27岁的纵贯追踪样本调查,梅纳德等人发现,在青春期结婚的人,滥用药物和精神健康问题呈正相关(Menard et al., 1989),但青年时(21岁至27岁)结婚的人,滥用药物和精神健康问题则呈负相关。青少年时期,在校读书与非法行为、滥用药物和心理健康呈负相关。沃福德分析同一样本,发现就业率与青少年严重罪行的犯罪率的相关性较高,但与青年期严重罪行的犯罪率相关性较低(研究参加者是18岁至24岁)(Wofford, 1989)。实际上,这些结果需要解释。从生命历程的角度看,可能特定年龄具有特定的行为规范(学校、婚姻、工作),违反这些准则可能增加参与非法或问题行为的风险。在研究方法上,这些结果表示,变量间的关系随生命历程而改变,也适合去测试这种变化是否存在。从截面数据来看,这种差异可能由于年龄或世代间差异;从多世代的纵贯数据来看,这些差异在某种程度上可能是年龄上的增长,而不是时期或世代差异。

因果关系[2]

有3个准则可用于检验各对变量之间存在的因果关系(Asher, 1983; Babbie, 2001:75—76; Baltes & Nesselroade, 1979: 35; Blalock, 1964):(1)问题中的现象或变量必须共变,例如,实验组和控制组之间的差异或两个变量之间的非零相关。(2)关系一定不能归因于任何一个或一组变量,也就是说,它绝不是虚假的,即使控制某些变量,关系仍存在。例如,实验设计中准确的随机分配(实验组与控制组干预前没有差异),或当其他变量保持不变时,两个变量之间的偏相关性不为0。(3)假定在时间上,因比果必须先或同时出现,即因的变化不会比果更晚出现。[3]第一和第二个准则可以用纯粹的截面或时间序列截面数据证明。第三个准则通常要纵贯数据才能充分地检验。但有一个例外,如果是变量的生物或遗传特征(性别、种族)所产生的影响,我们就能安全地假定这类变量没有纵贯数据的时间序列,因为实际上,当固定的特点被认为是某一变量特质(政治态度、非法行为)的原因时,我们至少具有部分时间序列数据,我们知道固定的特点必须先出现。换言之,在某特定时期测量固定的生物或基因特

征,但这结果从出生开始就是永存的。

如果出现非递归因果关系,情况就变得更加复杂。有些理论认为,因果影响不单从 X 到 Y,也可以从 Y 到 X。例如,马尔萨斯(Malthus)假定:(1)增加人均粮食供应会提高生育率;(2)提高生育率减少了人均粮食供应(Appleman, 1976)。如果事先没有一致的方案去消除因果关系的方向,只有截面数据就不能解决二元因果关系的顺序或方向的问题(Blalock, 1962; Heise, 1975; Simon, 1954)。当遇到非递归的因果模型时,会出现一个负反馈的回路,如马尔萨斯理论所建议的,只有截面数据难以充分地模拟实际过程,就算使用二阶段最小二乘法(Berry, 1984)、结构方程模型(Bollen, 1989; Hayduk, 1987; Kaplan, 2000)或其他比较复杂的数据分析方法也不行。纵贯数据(重复截面数据并不足够)比较容易解决因果次序的问题,可以测验因果影响的两个方向(但逻辑上不能完全得到保障,用纵贯数据来估计相互影响会更可靠,例如,测量时期可能无法精确地区分某一变量与另一变量出现变化的时间,从而解决这个难以辨认因果次序的问题)。马尔萨斯理论的例子中的负反馈回路可能需要相当长的时间序列数据才能充分地检验这一理论。

阶段性时间序列分析与定性变量的因果次序

在某些情况下,可能可以分辨两个变量的"启动"并确定两者的真正时间顺序。这样,就不必去推断第一个变量的改变导致第二个变量的改变(必须仍符合共变和非假的假设),但这种测试证明,第二个变量的变化不会引起第一个变量的变化。当变量被编码以表示假定起因变量或假定因变量(效果)是否已变了之后,阶段性分析就可用于确定这些变化的时间次序。这些变化可用简单的二分法来衡量(有,即变化已经发生;没有,即变化没有发生)。另一个重要的变化种类就是首次出现某一种行为的状况或种类。这指个案第一次进入某特定状况,或相应的,个人第一次出现某行为的特征。其他可能的变化包括行为的升级(进入序级中一个更高境界)或降级/减轻(进入较低的情况)以及停止(永久或暂时没有出现某种行为)。

犯罪学家提出 3 个假设:(1)滥用药物导致其他非法行为;(2)其他非法行为

导致滥用药物；(3)滥用药物和犯罪的原因是相同(例如，薄弱的传统道德观念、结交违法或犯罪朋友)。赫伊津哈等人的研究把每个受访者在每个时期有没有非法行为进行编码——“从来没有”或“曾经有”(即使当时已经没有)(Huizinga et al., 1989)。如果其他行为还没出现，出现过的行为算是先于其他行为。赫伊津哈等人发现，开始滥用药物(包括喝酒)通常跟随其他非法行为出现(证据反驳滥用药物会导致其他罪行的假设)。对于这些能确定时间序列的受访者(不包括在这研究之前或同一年出现两种行为的人)，所有受访者中曾经有非法行为(不包括酒精、大麻、硬毒品)和饮酒这两种行为的人，都是先有非法行为；所有曾经涉及非法行为和吸食大麻这两种行为的人也先有非法行为；所有曾经涉及非法行为和使用硬毒品这两种行为的人，也都先有非法行为。如果原因必须早于后果出现，那么非法行为就是饮酒和吸毒的起因。最可信的结论是，非法行为导致饮酒和滥用药物，或非法行为和滥用药物都有共同的起因(即关系是虚假的)，非法行为往往先于饮酒或滥用药物发生，可视为起因。埃利奥特及其同事的研究结果显示，后者的解释(虚假的关系)更有可能(Elliott et al., 1985; Elliott et al., 1989)。在一项相关研究中，梅纳德和埃利奥特测试了两种理论，一个是结交违法朋友会导致犯罪行为，另一个是犯罪行为导致结交违法朋友(Menard & Elliott, 1990a)。他们发现，开始结交违法朋友通常先于犯罪行为，这说明第一种理论(学习理论)成立，第二种理论(控制论)不成立。再者，虽然时间序列并不能充分地表示因果关系，但是它确实提供了证据以支持一个可信的因果关系。

用阶段性分析来检验时间或因果顺序并不总是可行。在某些情况下，调查过程或关系需要一段很长的时间，也不可能收集到刚开始的数据。这就是左侧删失的问题，即无法检测到刚出现的变化，因为它在收集第一期数据时已经发生了。此外，用来分析的变量可能具有无意义的开始或中止。在多国模型中，例如，没有一个国家的经济生产曾经是“零”，或“零”死亡率，或生育率，这种每个国家都有的持久特征，不能够通过提问谁先出现而去建立因果或时间的顺序。两个变量的变化是从一个时期到下一个时期，阶段模型不能分清哪个是因，哪个是果。

测量时间序列、因果次序和线性追踪样本分析

无论哪一个变量先于另一个变量变化，如果“开始”是没有意义的，那么单用变化前后发生的时间是不能决定这两个变量的时间序列或因果关系的。在第 3 章中将会更详细地解释，尤其当这些分析混合了点和区间的变量时。点测量是指某时间点（例如，面试的日期）。态度测量通常是点测验。区间测量涉及事件的计算、频率或一段时间的测量（例如，面试的前一年）。许多行为，特别是频率（多少次）和时间跨度（多久）的测量，都是区间测量（不能与区间尺度混淆，这是尺度的特性；区间测量的定义是测量时间的长短或次数）。事实上，点测量需要一个很短的时间跨度（现在），区间测量需要一个很长的时间跨度（去年一整年），但是这并不意味着有效的点测量只是一天或有效的区间测量是全年。这是完全可能的，例如，道德观念的测量从过去 10 个月到访问当天应该相当稳定（改变是从强到弱），而吸食大麻在过去 8 个月发生（先于那些没有吸食的受访者）。虽然测量的次数可能表示吸食大麻先于观念改变，但这个例子的真正时间序列（和因果）是观念改变先于吸食大麻。

一种可以确定因果方向并帮助解开那些具有无意义开始或结束的变量的因果次序（不用阶段分析）的方法，就是线性追踪样本分析（Finkel，1995；Kessler & Greenberg，1981）。在线性追踪样本分析中，我们想确定变量间的因果次序，这些变量都被视为因（内生）变量，并会测量至少两个时期（波）。测量时间相隔的长度称为“测量区间”（不要与“定距尺度”或“定距测量”混淆，后者请看前文）。模型中的变量可能仅被看做独立（外部）变量而只衡量一次，同时或早于第一个波中的内生变量的衡量。不管是否还有其他预测变量，至少每个内生变量有一个数值（滞后内生变量）可以当做该变量的最新预测值。方程内含有滞后内生变量有助于控制不可测量变量的影响，同时提供相对保守的非虚假、非零因果关系的测验。

线性追踪样本模型可包括任何一个滞后或实时影响，或同时包括两个影响。如果测量一个变量对另一个变量的影响时间等于或长于测量区间，这个效果便是滞后，可能多于一个测量区间；如果时间短于测量区间，这个效果便是实

时的。在线性追踪样本和阶段模型中，如果两个变量同时变化，那么测量区间可能太长，可以减短测量区间的长度，这使我们能够分开这两个转变。如果指定的滞后太短，即使强大的关系也可能变弱，因为全部自变量变化的影响尚未从因变量中反映出来。当测量间隔（或有些综合多元）不会紧随时间长度而变，它会等到全部自变量的影响都反映在因变量上，因此会有出现不正确的因果推论的可能性，或者更可能的是，低估了自变量对因变量的影响。

一个具有两个变量、三波的追踪样本模型如图 2.1 所示。变量是 X 和 Y，下标是指测量的时间或波。

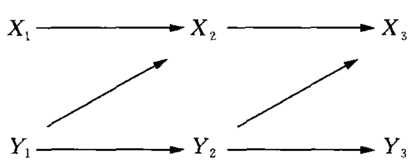


图 2.1　X 和 Y 的假定因果关系

图中的箭头表示，X 是最近的数值，该数值受下一个最近值（箭头从 X_1 到 X_2，从 X_2 到 X_3），但不是 X 前几个的数值的影响（没有直接从 X_1 到 X_3 的箭头）。同样的模型适用于 Y。此外，从 Y 到 X 有滞后效应（箭头从 Y_1 到 X_2，从 Y_2 到 X_3），这表示 Y 导致 X 的变化。没有任何箭头从 X 到 Y 表示 X 不能导致 Y。同时，图 2.1 表示，没有任何实时效应（X_1 和 Y_1、X_2 和 Y_2、X_3 和 Y_3 都没有箭头）。

梅纳德和埃利奥特用线性追踪样本分析来测试犯罪行为、对犯罪的态度和结交违法朋友的相互影响（Menard & Elliott, 1994）。使用三波模型，以前的行为、态度和关系模式都有可能影响后来的行为、态度和关系模式以及研究递归（无实时相互效应）和非递归（包括同步相互效应）的模型，他们发现，行为和关系模式同时相互影响。此外，他们发现：(1)态度影响关系模式，但不像关系模式影响态度这么强；(2)态度和行为互相影响，但很弱；(3)关系模式影响行为比行为影响关系模式大。开始出现犯法行为与结交违法朋友的频率的结果反映了之前描述的情况（Menard & Elliott, 1990a），但这里我们所关心的是每年犯罪行为的次数，而不是开始时间。这个例子说明了，当阶段分析和线性追踪样本分

析都可行时，结果可以互补，这也可能有助于发现复杂的互动或相互关系，但如不进行全面分析会很难发现。当有很强的理论支持一个确定的因果关系存在时，理论验证可以确认该假设时序或因果次序是否存在。当竞争理论指出不同和相反的因果次序时，时间序列分析或因果次序可提供一个更强烈的检验。

格兰杰因果关系

另一种测试因果方向和强度的方法是格兰杰因果关系（Cromwell et al.，1994）。X_t 和 Y'_t 两个变量都可以用平稳时间序列来表达（见第 5 章），其平均数为 0，

$$X_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + e_t$$

$$Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + f_t$$

e_t 和 f_t 是不相关的"白噪音"（误差），M 大于 0 但小于时间序列长度。根据格兰杰因果关系的准则，如果部分 b_j 不等于 0（必须是显著的），那么 Y 导致 X。相应地，如果部分 c_j 不等于 0，那么 X 导致 Y。实际上，格兰杰因果关系检验的问题是："有没有一个变量的变化不能用过去的数值来解释，但可以用另一个变量过去的数值来解释?"如果答案是肯定的，那么第二个变量就是第一个变量的"格兰杰起因"。请注意，如果 $m=1$，同时只有两个时期，格兰杰因果检验会降低到检验外生变量系数（所有都在时间 1 测量）对内生变量（在时间 2 测量）的显著性。当滞后内生变量（在时间 1 测量）也包含在方程内时，Y_2 模型就是 Y_1 和 X_1 的函数（时间 1 有可能不仅测量一个 X 变量）。模型不包括实时效应（例如，从 X_2 到 Y_2），从纵贯重访样本，这相当于不包含实时效应的两波线性重访模型。

m 是任意选择的，但必须在时间序列之内。巴纳德和克劳特曼（Barnard & Krautmann，1988）用单一的滞后内生变量（Y_{t-1}）与 X，X 测量了 3 个时期（X_{t-1}、X_{t-2}、X_{t-3}）。作为模型 Y 的起因，X 必须能解释 Y 过去或未来值都无法解释的变异。赖特（1989）分别分析内生变量滞后 1、滞后 2、滞后 3、滞后 4 和滞后 5，不同滞后有不同的结果。例如，格兰杰因果关系可以确认滞后 3 和滞后

4，但不能确认滞后 1、滞后 2 或滞后 5。我们如何解释这样的结果？

一般来说，方程内包含内生变量越早的数值，越有可能拒绝格兰杰因果关系的假定，但当它加入内生变量的附加值超过某数量时，可能就没有显著的影响。这个数量的估计是通过模型内生变量为自回归时间序列，或加入一个附加的滞后内生变量（例如，加上 Y_{t-4}），分开计算普通最小二乘回归模型，然后检测解释方差（R^2）的变化，如果解释方差没有显著性（Agresti & Finlay, 1997），那么加入该项变量似乎毫无意义。对于解释因变量的方差和拒绝格兰杰因果关系的假定，滞后 1 已经足够。这可能是因为 X 和 Y 不是平稳的时间序列。短时间序列可能不是很大的问题，但对于很长的时间序列，应该测试其平稳性。对于中等时序，可能无法确定是否适用于格兰杰检验。研究中似乎有滥用测试以及不能视平稳性的假设为理所当然的可能性。应特别注意应用自回归综合移动平均（ARIMA）时间序列方法来分析太短的时间序列，第 5 章会更详细地讨论该问题。

因果分析的其他问题

除了因果次序和存在相互影响，纵贯数据和因果模型分析可以用来区分行为长期和短期的影响。麦科德发现，童年攻击性的行为、父母打骂、控制和感情长期以来会形成侵略行为（McCord, 1983）。佩里学前教育（Perry Preschool Project）的一项有关学前启蒙计划的研究（Berrueta-Clement et al., 1984; Schweinhart & Weikart, 1980; Weikart et al., 1978）发现，这种启蒙教育对学生行为和学习仅存在短期效应，因为该效应似乎在随后几年就消失了，然后当学生进入青春期时，影响重新表现为长期影响。如果用历史事件分析方法（Blossfeld et al., 1989），只要有适当的数据，分析就更容易，可直接将年龄、时期效应和因果分析结合起来，以便解释发展和历史性的变化。这些例子属于研究变化和因果分析这类广泛的议题，但他们指出，有关变化和因果关系的问题（长期对短期的变化，对变动率的因果影响）有时比简单的问题（是否和为什么发生变化）更复杂。这类问题对多个纵贯设计可能会有重要的意义。

纵贯数据的巧合性和意图性

最早的社会科学的纵贯数据是全国人口普查数据，最初收集的目的可能不是测量改变或建立因果关系的方向或强度。早期普查的两个目的就是征兵和税收(Thomlinson, 1976)。之后，如何在宪法上规定美国各州直接税的分配便成了普查的目的，最近几年的普查也是用来作为联邦政府向各州分配资金的依据。事实上，普查数据可以用来衡量变化，最近也用于推断因果关系的性质，这主要出于偶然性而不是有意而为的结果。许多其他纵贯数据也有同样的现象。在 20 世纪，特别在第二次世界大战以后，研究者便开始特意收集纵贯数据以分析变化和因果关系，一般都是社会科学研究，特别是纵贯研究。尽管如此，研究中用到的个案大多是二手数据，不是一手数据，原因是政府机构都会定期收集数据。因此，不同时期的数据收集方法或变量定义可能会改变，因此很难得到完全可比较的数据。

在人口学上，联合国试图设立数个婴儿死亡和计算婴儿死亡率的标准。除了跨国家标准化的问题，采取这些标准的国家，如 1960 年的瑞典和 1975 年的西班牙(Hartford, 1984)得出了与以往不连续的数据，因此，比较 1970 年至 1980 年西班牙的婴儿死亡率便会出现不精确的问题。完整的记录或事件的计算在不同时期可能有所不同，那么变化的粗测量就会有问题。总统执法和司法委员会察觉到，由联邦调查局统计的 1958 年之前的美国农村犯罪记录数据“不完整或不可靠”，因为这是几年后才收集的。这种问题一开始就可以利用纵贯研究来处理，但要注意的是，因为这些数据最初有其他用途，所以研究人员应认真审核，以确定数据是否适合。变化单位或事件是否都已定义和计算？个案样本是否足够？有些数据可能不太适用于纵贯研究。

第 3 章 | 纵贯数据收集设计

不完全纵贯设计

第 1 章比较了纵贯研究和截面研究，截面研究的每个变量、每个个案仅在同一时间收集一次数据，还描述了前瞻追踪样本、回顾追踪样本和重复截面设计。但有些研究不能清晰地定义为纵贯研究或截面研究。阿鲁瓦利亚使用人均国民生产总值和收入不平等数据来检验收入不平等和经济发展之间的关系(Ahluwalia, 1974、1976)。由于收入数据不平等是零星收集的，阿鲁瓦利亚的研究使用的收入不平等和人均国民生产总值的数据是在同一时间测量的，但不同国家(个案)在不同的年份收集数据(例如，有些国家在 1955 年测量了某两个变量，但另外一些国家却在 1972 年才测量这两个变量)。一个单一截面分析，其数据收集的时间横跨了 18 年(1955 年至 1972 年)，阿鲁瓦利亚的分析假定这 18 年为一个时期。同样，这种方法假定人均国民生产总值和收入不平等是稳定的(数值没有重大变化，或者至少在国家间的排名变化不大)，或者至少这些变量之间的关系在这 18 年内基本不变。作出这样的假设前应该深思熟虑，最好有实验研究结果的支持。与稳定假设相反，梅纳德证明，收入不平等会随着时间而改变，甚至包括阿鲁瓦利亚的数据，这些国家收入不平等的排名也一直在变(Menard, 1983、1986)，但人均国民生产总值保持稳定。实际上，阿鲁瓦利亚的数据代表了一系列的截面数据，即重复衡量同一变量，但案件和时期都不相同。由于时间横跨太长，我们有理由怀疑把所有案件当做同期测量的适当性。

还有些略有不同的方案，托尔纳伊和克里斯滕森故意选了不同时间测量的

一些变量，以分析生育率、家庭计划与发展的因果路线(Tolnay & Christenson, 1984)。在所有国家和同一时间对每个变量进行测量，但不同的变量在不同的时间测量，这是为了配合路径模型中时间序列和因果次序。这与阿鲁瓦利亚的研究相反。阿鲁瓦利亚的每个个案的变量都在同一时间测量，但个案却在不同的时间测量；托尔纳伊和克里斯滕森刚好相反，每个变量的个案都在同一时间测量，但变量却在不同的时间测量。虽然不同变量在不同时期测量，但是每个变量对每个个案只测量一次，数据不能用于纵贯分析(例如，衡量变量从一个时期到另一个的变化)。托尔纳伊和克里斯滕森的研究本质上是截面设计。他们认为，如果有实时效应，就可以像纯粹截面数据那样进行分析。为了研究目的(评估家庭计划与发展对生育率的直接影响和间接影响)，这种设计是合适的，而且比路径模型的因果次序和测量时序都不同的模型更好(Menard & Elliott, 1990a)。托尔纳和伊克里斯滕森的设计具有时间序列数据和截面分析，可以当做一个时间序列截面设计。虽然按原先的定义，它不是真正的纵贯设计，但对于分析因果关系，它优于纯粹的截面设计。

应用时间序列截面数据，最好先确立一次时序，但难以确保可用果“预测”它的因。假设变量 X 和 Y 真正的因果关系如图 2.1 中所示：X 的前值影响 X 的后值，Y 的前值影响 Y 的后值，Y 的前值影响 X 的后值。这就是在特定有限的时间内，Y 是 X 效应的因。假设我们误认为 X 是 Y 的因，并在时间序列截面设计中包括了 X_2 是 Y_3 的因，而排除了所有其他 X(X_1 和 X_2)和 Y(Y_1 和 Y_2)。尽管因果关系是错误的，我们也可能发现 X_2 和 Y_3 之间的关系，如果 X_2 直接受 Y_1、Y_3 的影响而间接受 Y_1(通过 Y_2)的影响，那么 X_2 和 Y_3 就有一种虚假的关系。如果变量随时间相对缓慢地变化(即它们相对稳定)，那么当我们比较 X_2 和 Y_3、X_3 和 Y_2，或两个变量的纯粹截面组合(例如 X_2 和 Y_2)时，会得到相似的相关性。对于一个真正的纵贯设计(例如，前瞻重访设计)和分析，可以确定 X 和 Y 之间的真正因果方向。使用截面数据，尤其是时间序列截面数据，会有错误设定的风险，因为可能错误地估计模型的因果次序。纵贯数据比较容易侦测到不正确的因果关系。

总人口设计

图 3.1 显示了 4 种类型纵贯设计。在图 3.1 中,水平方向代表数据收集的时期,垂直方向代表收集的个案。在总人口设计中,所有人都会同时被调查或测量。因为每个时期都会有死亡和新生的人,所以不同时期的个案会有所不同。但如果时间很短,绝大部分的个案会相同。例如,美国 10 年一度的人口普查收集每 10 年的美国总人口的年龄、性别、种族和居住地,精准度估计为 95% 至 99%(Armas, 2001; Hogan & Robinson, 2000; Robey, 1989)。有些较少但又相当准确和完整的数据,如联邦调查局《统一犯罪报告》的数据,收集了各种罪案的被捕数据,对于某些罪行,美国各警区有罪犯的年龄、性别、种族与居住地(城市、郊区或农村)。

无论如何收集数据,总人口设计可能会有缺失数据或测量误差。

因为包括总人口,该设计应该适用于测量或推断时间趋势,但要密切检查年龄和世代效应(如早前讨论过)以清楚地确认趋势的性质。例如,被捕人数或比例的变化反映人口构成的变化(反叛青春期人数的百分比)多于反映个人或组群的行为变化(Chilton & Spielberger, 1971)。总人口设计与其他设计同样有年龄、时期、世代的独立效应的问题,但没有其他特殊问题。发展性变化可以利用截面研究(特定年份、跨世代)和纵贯研究(特定世代、跨年份,如果有足够的时间去检验发展变化)来检验,由此可比较两种方法所得出的发展效应的结果。适当选择时期、世代或者总人口,任何类型的纵贯数据收集方法都可应用在总人口设计上,但要有足够的独立时段以配合收集的方法。例如,线性追踪样本分析通常需要 2 期或 3 期(Finkel, 1995; Kessler & Greenberg, 1981),但 ARIMA 时间序列模型需要涉及 50 个不同的时期(Box & Jenkins, 1970:18)。如果测试模型,可能需要 100 期至 250 期才有充足的说服力(Yaffee & McGee, 2000)。

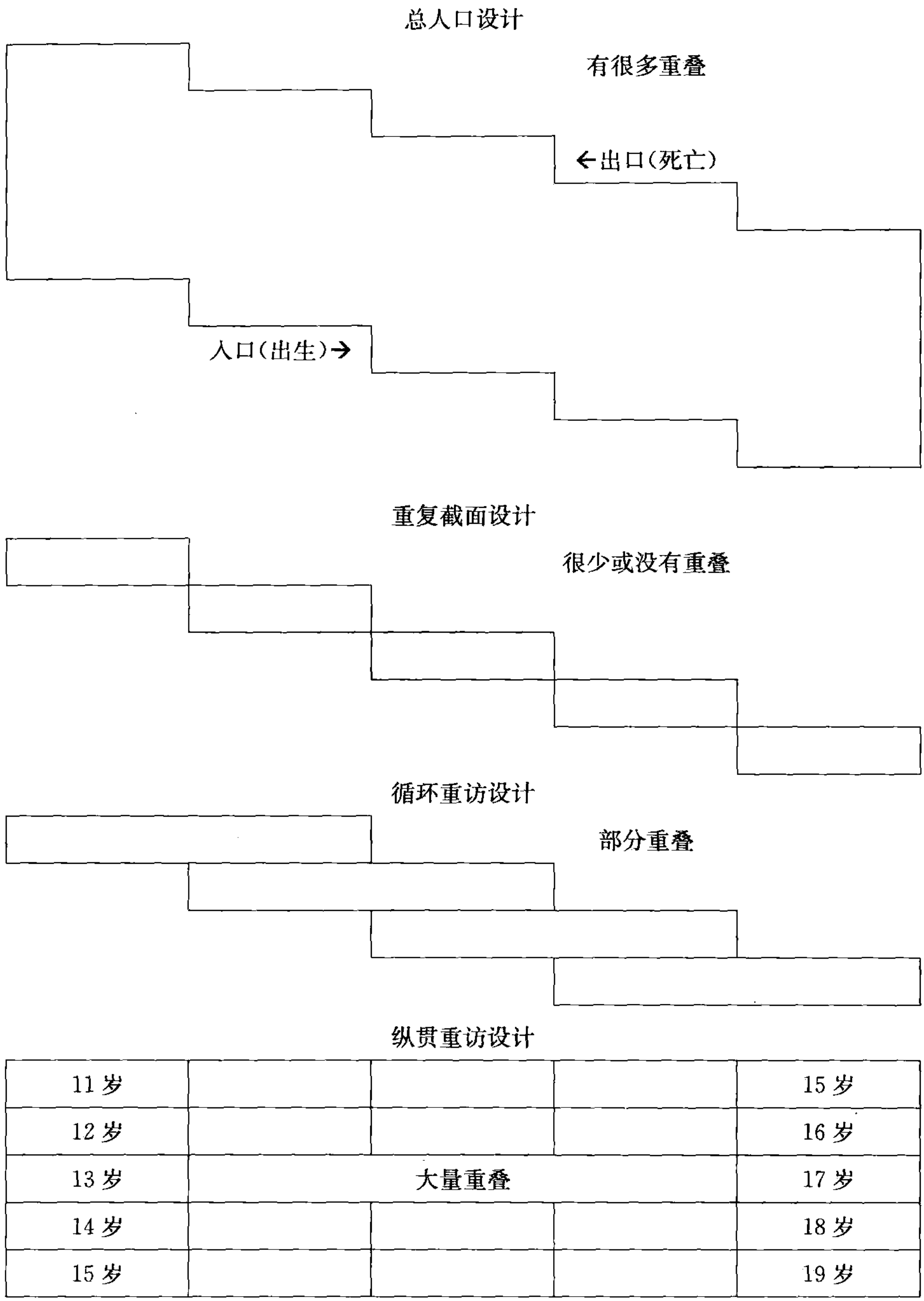


图 3.1 收集数据的纵贯设计

图3.1中的3个纵贯设计使用从总人口中抽取的样本，因此是总人口的一个子集。这3个设计在某种程度上在不同时期都具有相同或可比较的个案。这种区别对于在不同的研究中应用不同的纵贯设计是相当重要的。

重复截面设计

在重复截面设计中，研究者通常在每个测量时期选取不同的样本。因此，各个时期都包含完全不同的样本，就算重叠也会非常小，小到可以忽略，但不同时期的个案可以进行比较，因为它们都来自同一个总人口设计。重复截面设计以马丁·高德(Martin Gold)和他的同事做的全国青年调查(Gold & Reimer, 1975; Williams & Gold, 1972)为例。高德和他的同事收集了两个独立的全国青年概率样本，分别是1967年和1972年的调查。从这些样本中，他们推断，虽然青少年犯罪逮捕率从1967年至1972年都有变化，但该期间的自我报告犯罪行为却没有实质变化。约翰斯顿等人的"监测未来研究"自1975年以来，每年都收集全国高中生概率抽样。这些重复截面数据，如高德和他的同事的研究数据，可以分析各特定年龄组的态度和行为随着时间的变化趋势。"世界价值观调查"(Inglehart, 1997)可用于检查态度和信念、经济发展和政治文化之间的关系以及该关系的长期变化。其他重复截面设计的例子包括民意调查、选举投票和一般社会调查，美国的全国民意研究中心每年的总人口调查的议题范围广泛，包括婚姻和家庭、性行为和性别角色、劳动力参与、教育、收入、宗教、政治、犯罪和暴力、健康和个人幸福，并强调问题能准确地被复制，以促进跨年代的比较研究(Davis & Smith, 1992)。

重复截面设计的主要限制是难以分析世代间的发展模式，并难以分析因果次序。这两种限制源于重复截面的设计方法，相同个案不会重复或在不同时期重复测量。发展模式的研究是通过观察测量不同时期、不同年龄的差异(即跨世代)，重复截面设计相比纯粹截面设计的唯一好处是，重复截面设计可能得出不同时期的截面结果。这将减少但不能完全消除跨世代的发展模式无法反映世代内发展模式的可能性。对于因果次序，个案具有两次或两次以上的测量的

缺失数据意味着，不可能使用阶段和线性追踪样本分析（除非有充分程序去匹配不同时期的不同个案，不过这点不太可能）。重复截面设计中的变化测量只可用于样本或子样本这种集体数据，如男性和女性、民族或社会阶层，它不能用于个人层面的样本。这样的限制使许多研究人员——尤其在发展心理学方面（例如，Baltes & Nesselroade, 1979）——认为，重复截面设计不算纵贯设计。然而巴尔特斯等人的世代效应研究却应用了重复截面设计（他们称之为"截面序列"）。一般而言，重复截面设计适用于测量集体时期趋势。如果已经知道因果次序，而且因果之间的时间滞后相对于测量相隔时间要短，重复截面设计就可用来分析模型的因果，该模型的性质基本上是截面的。如果世代间和世代内的发展差异密切地影响彼此，那么使用多世代的重复截面设计来分析截面的发展模式也没有问题。对因果推论和发展分析，研究者需要应用其他纵贯设计以肯定重复截面设计是否合适。最后，重复截面设计能复制不同时期的截面结果。如果我们想研究世代内的发展变化或确定因果次序，应用其他纵贯数据的收集方法会比较好。

循环追踪样本设计

循环追踪样本设计的数据收集方法如下：不管是回顾测量样本还是前瞻测量样本，都取部分测量时期，然后减少一些样本，再加入新的样本。循环追踪样本设计可以降低前瞻式研究的小组死亡率和反复测量的问题（将在第 4 章讨论）或回顾式研究的回忆时期的问题。在几个测量时期保留某些个案可以分析个人层面的改变、世代内短期的发展变化和小组。样本的更换、加入新的但可比较的子样本能够分析总体变化的长期模式。如果因果关系的滞后时间比留下的个案时间短，那么就有可能分析时间序列和因果次序。纵贯数据包括了部分不同个案的不同时期的重复测量数据，研究者可以比较重复测量是否会产生任何数据上的偏差（例如，建立某种程度上的信任，会更愿意或不愿意回答问题，或觉得跟进问题太冗长）。

美国全国犯罪调查[4]由司法部赞助、美国统计局负责调查，就是一个很好

的循环追踪样本设计例子。该调查组织定期访问住户成员家人的犯罪受害，包括强奸、抢劫、严重殴打、侵犯他人、入室盗窃、偷窃和偷车等7种罪行，并利用概率抽样选取受访家庭，3年内访问过7次（每6个月一次）的家庭会被新抽选的住户所取代。家庭作为单位分析，这能够分析短期家庭内受害率的趋势、整体或平均受害率的短期和长期趋势，但是不能分析家庭内的长期发展趋势。

堪萨斯市警方巡逻实验（Kelling et al.，1974）也使用循环追踪样本设计来收集受害者的数据。前测访问包括了1200户家庭，后测只保留一半，另一半被新样本取代了。因此，可比较那些被访问了两次（前测和后测）与一次（仅后测）的人，同时可以排除重复访问可能带来的偏差（控制组和实验组没有差别）。

循环追踪样本设计适用于研究个人在特定年龄范围时的情况，如青少年或65岁以上的老年人，这样可以防止过了青春期或死亡所带来的严重的样本问题。某年龄范围的循环样本可以让研究人员保持足够的个案数目（参阅Kraemer & Thiemann，1987），以进行更复杂的分析或子样本分析。

纵贯追踪样本设计

在纵贯追踪样本设计中，研究者每段时期都访问同样的个案。在实践中，每期可能会有点变化，会产生遗漏数据。例如，当个案以个人为主时，研究期间有些人可能会死亡，有些可能不想再参加，有些移居到别处，有些个案的研究者无法找到。这些都是流失样本的主要原因，特别是需要纵贯几次的前瞻式追踪样本设计。样本的流失对回顾式追踪样本设计而言不是问题，研究者可能每个时期收集一次数据，然后收集两个或两个以上的时期（在数据收集期间或之前）。在回顾式追踪样本设计中，可能有样本偏差，因为它排除了最后数据收集期间已死亡的受访者，或者有前期但没有后期数据的人。回顾式和与前瞻式追踪样本设计都会有遗漏数据，因为受访者可能记不清过去的事件、行为或态度，或不愿透露某些资料，还有研究者无法找到受访者或受访者不愿合作的可能性。原则上，这两种设计的数据质量应该没有分别，但实际上是有的。

纵贯追踪样本设计包括多个世代（如图3.1），如果具有足够的时期和世代

数目，就能进行任何类型的纵贯分析。单一世代追踪样本设计不能比较不同的世代，但多世代设计就可以分析年龄、时期和世代效应，描述发展和历史变化，分析事件的时序，并可进行线性追踪样本分析和因果分析。以全国青年调查为例，埃利奥特和他的同事利用概率抽样在全国选取了 1976 年 11 岁至 17 岁（7 个世代）的青少年，同时对这些人进行定期访问，最近一次是在 1993 年（Elliott et al.，1985；Elliott et al.，1989）。全国青年调查的数据用以分析以下问题：(1)估计和分析非法行为的期间趋势（Menard，1987b）；(2)分开滥用药物（Elliott et al.，1989；Menard & Huizinga，1989）和非法行为（Elliott et al.，1989；Menard & Elliott，1990）中的年龄、时期及世代效应；(3)测试和复制青少年与青年的非法行为理论（Elliott et al.，1985；Elliott et al.，1989；Roitberg & Menard，1995）；(4)显示非法行为和其预测变量关系的发展变化（Menard et al.，1989；Wofford，1989）；(5)检测青春期到成年非法行为的持续性（Menard & Mihalic，2001）；(6)确定变量的时间次序，以解决不同竞争理论之间的矛盾（Menard & Elliott，1990a）。这些例子说明，多世代前瞻式追踪样本数据具有广泛的应用性。

其他变化

图 3.1 并非纵贯研究唯一的可能设计。例如，在循环样本中，有可能某些样本在某一个时期没有，但下一个时期又被重新纳入。另外，也有可能当个案到了某些标准就不能再包括了，但也没有再加入新样本，这样，样本数量会逐渐减少，之后几年的数据分析可能会有问题（除非修正设计，加入年轻的世代）。一般与数据收集设计相关的考虑都不会变，但是，修正图 3.1 的设计必须评估这些基本设计的变异，包括是否足以描述短期和长期的历史趋势（时期效应）？是否足以描述世代间或世代内的发展变化（年龄影响）？是否能分离年龄、时期和世代效应？是否能确定因果影响的强度与方向（例如利用阶段分析或线性追踪样本分析）？只要有足够的世代和测量时期，总人口设计和纵贯追踪样本设计几乎可用于任何类型的纵贯分析。应用其他的设计有一定的限制，研究者必须根

据特定的研究问题来判断其适用性。

由于每个设计需要的个案和时期的数量可能不一样，因此分析方法可能也不一样。如果个案和时期数量很大（例如几千个或更长时间），分析方法就要根据数据的质量而定。如果个案和时期数量很少（例如 1 个至 10 个或 2 个至 10 个时期），那么任何数量分析都可能有问题。[5]另外，如果个案数目颇大（例如，个人层面有 1000 个个案，或集体层面有 50 例），但时期数量很小，那么就可以应用线性追踪样本分析（Finkel，1995；Kessler & Greenberg，1981）。另外，个案的数量小及时期大的情况更适合用时间序列分析（Wei，1990；Yaffee & McGee，2000）。原则上来说，个案数目与设计类型无关。在总人口设计上，例如在个人层面，部落社会的总人口可能少于 100 人。综合分析的单位可能是世代或总人口而不是个别成员。全国犯罪调查的最后时期包括了 6 万个家庭，共有 10 万个人。纵贯研究很广泛，包括了数据收集设计和个案数目的所有可能组合。第 5 章会讨论个案和时期数量不同的组合和分析方法。

第 4 章 | 纵贯研究问题

纵贯研究并没有独特的收集数据方法。纵贯研究像截面研究一样，基于 3 个最基本的收集方法：提问、观察人们的行为、观察行为变化的痕迹或结果。不同的研究都需要收集数据，如单一个案、小数量样本、非常大数量的样本、社会上所有的人、社会中的概率样本或社会上特定组群的研究。数据需要保存和编码，不管是个人层面还是家庭、人口、国家等集体层面。数据收集的时期或短（几个小时）或长（几年），并需要制定标准的程序。纵贯或截面研究可能涉及个案分析、人种学、实验、样本调查、人口普查、档案数据，但纵贯与截面研究最主要的区别是收集数据，如前文所述，纵贯研究中的每个变量至少有两段时期的测量资料。

纵贯研究与截面研究同样要面对数据质量的问题。内在或外在测量效度、量表的可信度、抽选样本、提出适当的问题、实验设计的随机程序、受访者与实验者、访谈或观察人员的互动效应（小型社会数据收集）、研究相关的问题（我们是否测量到重点或只是一些容易测到的东西）和研究经费，这些无论在纵贯或截面研究中都是相当重要的。有些问题在纵贯研究中比在截面研究中更严重。重复截面设计的重复会增加样本偏差。相关讨论请看巴比、别克曼和罗格的著作（Babbie，2001；Bickman & Rog，1998）。这里，我们只集中讨论纵贯研究的问题，而非截面研究。

起源与预测

沙首（Zazzo）认为，应将研究分为行为的起源和对行为的预测两种（引自

Wall & Williams，1970)。在沙首看来，预测是考虑人口特征在一段时间内的稳定性和外在因素(环境变化、治疗干预)如何影响这些特征，相反，起源主要集中在质性变化的阶段和顺序，目的是发现成长或发展变化的规律。沙首认为，行为起源的研究方法更质性化：摒弃大样本、预定假设和变量；更趋向于小样本的深入研究；一开始没有预定假设哪个变量重要；摒弃年龄(时间序列)和会随着时间而变的连续测量。

不可否认，很多纵贯研究的重点在于预测而非起源(如沙首定义)，但无论在宏观或微观社会层面上，我们还是意识到起源的研究是不能忽略的。在宏观社会层面，罗斯涛提出一个经济发展阶段理论(Rostow，1960)，布莱克(Black，1996)尝试定义政治发展阶段或政治现代化以及人口转换理论(Caldwell，1976；Davis，1963；Notestein，1945；Thompson，1929)，虽然受到一些批评，但是该理论证明了人口学研究的时期理论和方法(Menard，1987a；Tolnay & Christenson，1984)。在微观层面，肯德尔和她的同事(Kandel，1975；Kandel & Faut，1975；Kandel & Logan，1984；Yamaguchi & Kandel，1984a、1984b)检试了滥用药物阶段顺序，结果发现从饮酒到吸食大麻到服用其他违法药品的过程中，很少有人一开始就使用最严重的药品。肯德尔的研究还显示了行为起源的研究不能完全与预测研究分开。她和她的同事不仅描述了药物滥用的过程，而且还解释了在怎样的情况下，人们会放弃一种药品而服用另一种。为了对行为发展模式有更完整的了解，我们可能需要研究行为发展次序和阶段转变的时间以及它们之间的相关性。

测量随着时间而变化

1930 年，雷德菲尔德发表了一份关于墨西哥村庄迪坡斯特兰(Tepoztlan)的人种研究(Redfield，1930)，其中他称那是一个和谐的村庄，人民善良又满足。21 年后，刘易斯也发表了一份同一个村庄的人种研究，与雷德菲尔德相反，他发现了该村和与其他村庄之间有相当严重的暴力冲突(Lewis，1951)。米德(Mead)在 1928 年发表了她对萨摩亚(Samoa)人种的研究。55 年后，弗里曼质

疑米德的研究,并提出与她相反的结论(Freeman, 1983)。因为时间的差异,所以没法判断结果上的不同是否反映了真实的变化,或只是观察者不同的偏见或取向而已。

这个问题也可能出现在调查研究中。马丁表示,重复截面设计无法复制美国机构的受害调查和信心调查。关于美国机构的信心调查,哈里斯(Harris)与全国民意研究中心做了几乎同样的截面研究,得出的结论和发展趋势都不一样。至于受害调查,城市学院(Urban Institute)根据统计局的全国犯罪调查所制定的标准,在全国犯罪调查的一年后,复制了同样的调查,结果得出的受害率比全国犯罪调查少一半,甚至比当年的电话调查还低。

第二个例子说明了潜在的测量问题,同样来自全国犯罪调查。1992 年,全国犯罪调查改变测量受害率的方法,也改名为"美国全国犯罪被害调查"(NCVS)。这次改变得出更高的估计犯罪受害率。我们分别将全国犯罪调查和全国犯罪被害调查的前一半或后一半的结果进行截面比较。1992 年后的数据需要做些调整以检验更长期的受害率趋势,但无法比较这两个系列的多年趋势。重新设计后,暴力受害在 1994 年达至高峰,之后急剧下降,但是财产犯罪持续下降,这早于重新设计之前就出现了(Rennison, 2000)。鉴于不一致的暴力和财产受害趋势,似乎不能相信暴力受害趋势的变化是由于测量设计的改变,但我们只看暴力受害,则有理由相信受害率自 1992 年后有所改变是由于测量设计不同,而非行为上的改变。

高德的违法行为重复截面研究(Gold & Reimer, 1975; Williams & Gold, 1972)成功地将第一波的抽样程序和重要的结果复制到第二波研究上。高德和他的同事的调查的其中一个优点是,主要研究人员是一致的。如果对研究程序和相关领域没有深入的了解,将难以或根本不可能复制前一波的数据收集,以完成重复截面或前瞻(总人口、循环追踪样本、纵贯追踪样本)纵贯研究。制定数据收集标准的问题就是,不同收集方法的标准会有不同,还需要问卷的配合,每轮都使用相同的调查问卷,但不同的研究人员也会产生变异,这也是一个问题。这个问题主要是访问员的训练。质性研究中收集数据量表的可能就是该观察员,不同的观察员有不同的偏差、观察和重点,导致不同的结论。有些研究

人员认为,参与观察研究的结果是难以复制的(Blalock & Blalok, 1982:97)。

缺乏标准化的数据收集可能会出现合理性的问题。如果纵贯调查青少年受访者,直到他们30岁左右,早期的调查重点可能在于学校,而后来可能是工作或职业。受访者会从原生家庭(父母和兄弟)过渡到再生家庭(配偶和子女)。从理论上来说,工作或学校的态度和家庭压力是预测某些行为(如滥用药品)的重要因素,因此受访者的行为改变可能有其相对应的生活变化。这里有两个重点问题:第一,变量的测量在某个人生阶段是否等同于后阶段中相近概念的另一个变量呢?例如,原生家庭的压力和再生家庭的压力。第二,过渡是突然的还是渐变的?受访者是否同时经历两件重要的事件(例如,工作同时又读书),或是完全离开一个然后再进行另一个呢?如果同时测量这两个内容,那么就有可能直接估计两者的关系,并了解它们与另一个变量的关系是否相似。如果变量是高相关的,而且如果它们与另一个变量的关系模式一样,那就有同时效度(Bohrnstedt, 1983; Zeller & Carmines, 1980)。

量度转变的另外一些原因是,研究本身或与科学社会相关的领域出现了新的假设,或者研究人员转变了他们的研究兴趣(Wall & Williams, 1970)。加入新的假设可能是转变的原因之一,不过可以明显看到其中的危险。如果其他研究明确地排除或反驳该研究所依赖的假设,那么继续研究就没有意义。完全驳斥一个理论或假设在社会科学中是很罕见的,但即使这样,数据仍可能复制出驳斥旧假设的结果。另外,在纵贯研究过程中改变假设、变量、测量,可能令研究前后两个部分的结果和原来所设想的东西无法比较,这可能破坏前后数据的实用性。另外,研究可能会受过渡性理论的影响。

追踪样本人数流失

在纽科姆和本特勒的青少年吸毒纵贯研究中(Newcomb & Bentler, 1988),8年间流失了55%的受访者。在默里和埃里克森的吸食大麻报告中,也有50%的流失率(Murry & Erickson, 1987)。其他研究却表现良好。克拉里奇等人访问了威斯康星高中学校高年级的学生(17岁),第一次访问后只有11%的

流失率(Clarridge, 1977)。登普斯特-麦克莱恩和摩恩在 1956 年访问了美国康奈尔大学的 427 位母亲并收集了 30 年后(1986 年)的访问数据(Dempster-McClain & Moen, 1998)。在这 427 位受访者中,4%(17 人)找不到,19%(82 人)已经过世,3%(13 人)拒绝参与跟进,73%完成再次访问。其他前瞻式纵贯追踪样本,如收入动态追踪调查(PSID)和收入与活动参与调查报告保留率(对于第一次访问人数)大约是 65%至 80%(Brown et al., 1996; Hill, 1999; Kalton et al., 1989)。全国青年调查(NYS)(Elliott et al., 1989)指出,5 年的纵贯研究的流失率低于 10%(每年访问),超过 17 年约 20%(访问相隔 1 年至 3 年)。但是结合最初的流失率,每年跟进调查访问的保留率也相当高,如 NYS 或 PSID,转化为整段时间的总损失率(至少遗漏一次的数据收集受访者的百分比)是 40%至 50%或总保留率 50%至 60%。

此后调查所失去的受访者可能混淆测量变化,因为他们可能与那些保留下来的受访者有系统上的分别(可能变量一开始就有不同的平均值,或者跟其他样本的改变方式不同)。如果不成比例流失,研究变量的极端值就会造成特别严重的误差,例如,最常见吸食非法毒品的人或非法行为研究中最严重的罪犯。因此,它不仅反映严重程度的流失,而且流失模式影响重要的变量,这是很大的问题,尤其是关于偏差或非法行为的研究。例如,沃尔顿等人发现,他们难以再次联络到治疗组滥用药物的受访者。科尔德雷和波尔克发现,即使有相对高的流失率,保留下来的受访者还是能够提供相对准确的二元和多元关系的估计,但对普及率和行为频率的估计就有偏差,特别是大众人口样本。索恩伯里等人发现,如果不包括独特的受访者,普及率、犯罪频率和毒品使用的估计可能会有偏差(Thornberry et al., 1993),但与科尔德雷和波克发现不一致的是,多元关系也会受到影响。

如果研究者不能与研究对象保持联系,流失率将不可避免地提高,伯吉斯、克拉里奇等人及登普斯特-麦克莱恩和摩恩详细地讨论了纵贯受访者的技术,包括在第一次和随后的访问中取得受访者父母、其他亲属、朋友或其他常联络的人的姓名和地址。如果受访者搬家,请要求邮局提供一个转发地址。电话簿、信贷机构、探访或登普斯特-麦克莱恩和摩恩所指出的越来越多的网上资源

也都可以利用(Burges, 1989; Clarridge et al., 1977; Dempster-McClain & Moen, 1998)。为了减少流失率,不管用什么方法,都要努力保留纵贯受访者数。

克拉里奇等人在17年里追踪了威斯康星高中超过1万名学生,保留了97.4%的受访者,成功访问了88.6%的学生(Clarridge et al., 1977)。克拉里奇等人使用各种方法,从受访者的父母、大学、高中、邮局、兵役、邻居和朋友,才获得如此高的回应率。伯吉斯总结,能够联络或追踪80%至90%的受访者是合理的,即使调查跨越时间很长(Burges, 1989)。

在某种程度上,受访者的流失会影响变量的分布和实际的结论。二元测试(Bulmer, 1979)可用于测试不同的人口种类(男或女、白人或非白人等)的人数比例是否随时间有显著的改变。其他显著性的分析可以检验第一轮数据收集后保留下来的受访者与流失的受访者在以下各方面是否有所不同:(1)对某些变量的数值;(2)变量关系(相关性)的强度;(3)3个或以上变量关系的结构(如多元回归方程或共变结构)。这些测试会揭开不同轮次的数据收集的样本变异。当然,它还是有可能无法检测到显著样本变异影响实际分析结果的某些因素。例如,不同的行为轨迹(如增加或减少滥用药物)可能与保留下来或流失的人有所不同,以上方法也可能没法侦察。因此,实质结果会有严重偏差,如行为的估计和发展趋势的解释,这很难观察到。

回顾追踪样本研究的流失率问题又是另一回事,不是显著性测试的评估,而是与离开第一轮数据收集的受访者相关,即选择问题。特别是在长期研究中,回顾追踪样本研究可能会遗漏某些人,如死亡或在抽样或调查期间离开该地区的人。这些人可能与其余人口有系统化的分别。例如,滥用非法药物的人比其他人有更高的死亡率。如果这样,那么研究期间滥用非法药物的人的抽样会被低估,而且研究会错误地估计滥用药物比率的变化。实际上这就是流失问题,但这在选样前已经发生了,所以它比前瞻追踪样本研究更难察觉和测量。

处理纵贯研究的遗漏数据[6]

多项量表的纵贯数据可能允许总体量表的数据上的单项问题的缺失数据，或某个案的某波段的数据（但不能是全部数据），或某个案的所有数据，例如第一次就没有应答的个案。不管哪种形式的缺失数据，都可能导致预测或因果模式或描述统计（如平均值、方差或趋势）估计上的偏差。如果多项量表具有大多数题目的数据，相对来说，它能较简单又准确地去填补该缺失数据，这可能比删除该缺失数据的个案更好。当整个量表的数据都遗漏了，有几个方法可以补救，虽然都不是很满意的做法，但至少比删除整个量表好。对第一次或某波段的缺失数据，如果该变量具有清楚的缺失模式（如种族或社会经济地位），通常的做法是加权该个案，这样可以减少潜在的误差，但这取决于该缺失个案与其他现存变量的共同特征（如种族或社会经济地位）和关系。

另外，可以通过模拟来调整回归或相似模型的系数，特别是那些不回应者与其他应答者有明显的分别。加权和模拟非应答的方法都有些不好的地方（Allison，2002；Brehm，1993）。一般来说，如果应答率低的话，加权会比模拟好，相反，如果应答率高（如超过 30%），则模拟比加权好，但对后者来说，两个方法都不令人满意。对波段缺失，可以用内推法（利用前后的数据来估计遗漏波段的数据）来填补一些方法。对那些变化比较少或有明确形式的变量来说，内推法或外推法在某些程度上都比较合理，但对经常变化的变量，如态度，就不适用了。

简单来说，没有任何一种处理缺失数据的方法是最好的。所有方法在缺失数据的形式或与现存数据的相似度上都有不可证明的假设。在某些情况下，比较简单的方法（例如，用量表中其他题目的平均值代替单项的缺失数据，或加权该个案以补偿流失者的观察或测量等特点，如种族或社会经济地位）都相当有效。在其他情况下，一定要选择简易而且更好地避免偏差的方法，例如，通过选择相对简单（而且广泛使用）的成列删除或更复杂的技术，如多重插补或回归插补来填补整个量表或某波段的个案。具体而言，研究人员可能会有 3 种选择：

(1)成为熟练的统计师,使用最大似然法或多重插补方法加选择模式;(2)聘请熟练的统计师和使用最大似然法或多重插补方法加选择模式;(3)使用成列删除。其实,选择(1)及选择(2)也可能使用成列删除和加权个案的方法。对于缺失个案,加权个案得到的结果往往能媲美很复杂的模型方法(Taris, 2000:35—36)。对于题目数据,方法上(如多重插补平均)来说比成列删除好些,但对于某特定数据集而言就不恰当。格雷厄姆和霍费尔建议,如果缺失个案少于5%,成列删除应可接受(Graham & Hofer, 2000)。艾利森也认为,相对于其他方法,包括最大似然法和多重插补(两者都基于数据缺失的随机性),特别是最小平方回归分析和 logistic 回归,成列删除比较不会违反数据缺失的随机性假设。但是,一般来说,最大似然法对估算刻度量表或非刻度量表的缺失数据比简单方法(如成列删除)好,至少当这些数据是随机缺失的且缺失比例很大的时候。

反复测量和追踪样本训练

追踪样本的条件习惯效应(例如,Kalton et al., 1989)对同群受访者的连续研究是一个问题,特别是微观社会的纵贯追踪样本设计,包括涉及前测的实验和准实验设计。反复测试的效应可能会损坏实验和伪实验的内部效度,但使用控制组能让研究人员测量这个效应,同时如果使效应存在,就可以决定是否有治疗效果和反复测验所带来的改变。全国犯罪调查研究的效度受到追踪样本访问的影响(Cantor, 1989)。梅斯奇和肯德尔发现,滥用药物研究也有相似的问题(Mensch & Kandel, 1998)。

受访者愿意回答问题的反应(例如跟进问题)是唯一影响同一个案的持续研究的有效性。不愿意参与研究的人可能导致跟进研究的流失率。另一种可能性是,受访者参与调查后,情况可能会改变。1984 年,全国青年调查进行的抑郁研究中,受访者若报告感到沮丧和有抑郁症状(根据临床上的抑郁量表),可要求匿名转介到专业精神健康部门。这个简单的选择改变了小部分受访者对该选项的态度或行为。柯林斯等人(Collins et al., 1989)在一项照顾家中老人的研究中发现,此改变对至少 52%的研究参与者有一项影响,最常见的是他们

(护理人员)如何应付在家照顾老人的限制(Collins et al.,1989)。鲁宾和米切尔报告,夫妻关系发展的纵贯研究也受到影响(Rubin & Mitchell, 1978)。共同的模式就是,受访者似乎对他们的态度、情感和行为更明白和内省。我们很难判断这样是否会令他们的态度、情绪或行为有重大改变,而且我们不清楚产生这些效应是否由于单一截面研究或在纵贯研究中与研究人员的反复接触。

不仅持续的调查研究才有这个问题,任何微观的社会研究,包括观察研究,如果其中的研究者与研究对象之间存在联系或研究对象知道他们正在被观察,都会有这种误差的风险。宏观社会研究是基于数据是如何收集或累积的,这个问题可能较少。美国的犯罪人口普查及生命统计数据的时间长度似乎没有受到任何负面数据效度的影响。如果有,相反就是个案。普查范围随着时间推移越来越全面(Robey, 1989)。1990 年的人口普查受到广泛的批评,因为它缺失了约总人口的 1.6%,其中遗漏了更多某人口群组的数据(7%的 18 岁以下的非裔美国人)。最初报告指出,2000 年人口普查也少了 1%,同样遗漏了非裔美国人(Armas, 2001; Hogan & Robinson, 2000)。美国联邦调查局统一犯罪报告的警区覆盖率同样随着时间而增加。国际人口统计数字、经济发展以及其他民族特色似乎没有随时间而变得更糟。请注意,回顾性研究可能不像其他持续研究那么容易受到反复访问问题的影响,尽管回顾性研究像前瞻性和截面研究那样,访问过程过于漫长而沉闷,可能也会出现问题。

持续研究中的追踪样本问题主要在微观社会和前瞻性的研究中,不过可以通过调整数据的收集时段(全国青年调查间隔 1 年至 3 年,全国犯罪调查相隔 6 个月),或改变不同时期的问卷设计(不过这可能有数据兼容性的问题),或成功地鼓励研究人员和研究对象高度投入该研究,这样可以避免这种问题。

受访者回忆

路特等人发表了一篇很好的针对短期与长期回顾数据的评论报告(Rutter et al.,1998)。短期与长期回顾数据通常显示:(1)记忆会随着时间淡去;(2)短期和长期回忆数据在突出的事件或态度方面更清晰;(3)短期和长期回忆数据

在客观事件或特征上往往比态度或其他心理数据更一致;(4)短期和长期回忆数据之间的分别表明,长期回忆数据与受访者目前的看法和态度往往会有一点偏差。对于最后一点,因为人们往往重建和重新解释他们的记忆,以反映其现时的生活情况和态度,从而建设一致的“生命故事”。这并不是说长期回忆数据不可靠或无效,不过,使用这类数据确实要谨慎,而且有可能的话,还是用前瞻研究的较短期回忆数据更好些。

当比较同一研究对象的前瞻(或者更准确、短期回忆)和回顾(或长期回忆)数据时,潜在后果就出现不一致。例如,亨利等人比较了档案记录的回顾数据和受访报告的前瞻数据,内容包括身体特点(身高、体重)、居住流动性、家庭关系、与警察的交流与联系、阅读能力和心理健康(Henry et al., 1994)。他们发现,心理或态度变量(如家庭关系)的回顾数据最不符合前瞻数据,而客观特征的测量(例如,居住改变或与警察的接触)相对比较符合。虽然回顾与前瞻数据的相关性很高,但是绝对差异有时很大。

索伦森等人认为,回顾设计能提供过去违法的准确数字(Sorenson et al., 1989),但是这个推论是基于比较两个不同人口的截面调查的(1965 年的康特拉科斯塔县以及 1981 年和 1982 年的圣路易斯市区)。基于 1965 年的圣路易斯截面数据(1981 年至 1982 年的回顾性收集)得出的非法行为结果与 1965 年的康特拉科斯塔县的结果相似,索伦森等人的结论是,圣路易斯回顾数据是有效的。该结论的缺陷是,这可能巧合。特别是康特拉科斯塔县 1965 年的犯罪率比 1965 年圣路易斯的犯罪率低,加上圣路易斯的回顾数据漏报了过去的非法行为,那么这两个地方的犯罪率可能相同,即使有回顾式数据效度的问题。更好的方法是将这些研究对象的前瞻和回顾数据进行比较。

梅纳德和埃利奥特利用全国青年调查的数据来比较:(1)基于 1 年、2 年和 3 年的回忆,犯罪普及的趋势(样本中受访者犯某项罪行的百分比);(2)基于 1 年回忆期的前瞻数据和变量的回顾数据,严重犯罪的普及性。图 4.1 说明了 1 年、2 年、3 年回忆期的非药物犯罪、严重(指数)非药物犯罪、吸食大麻和其他非法药物使用(使用多种药物)的比较。实线代表前瞻性的 1 年回忆(1976 年至 1983 年)(1981 年和 1982 年数据不详),虚线代表 1981 年的 3 年回顾数据的趋势和

1982 年的 2 年回顾数据(连接 1980 年和 1983 年的数据点)。

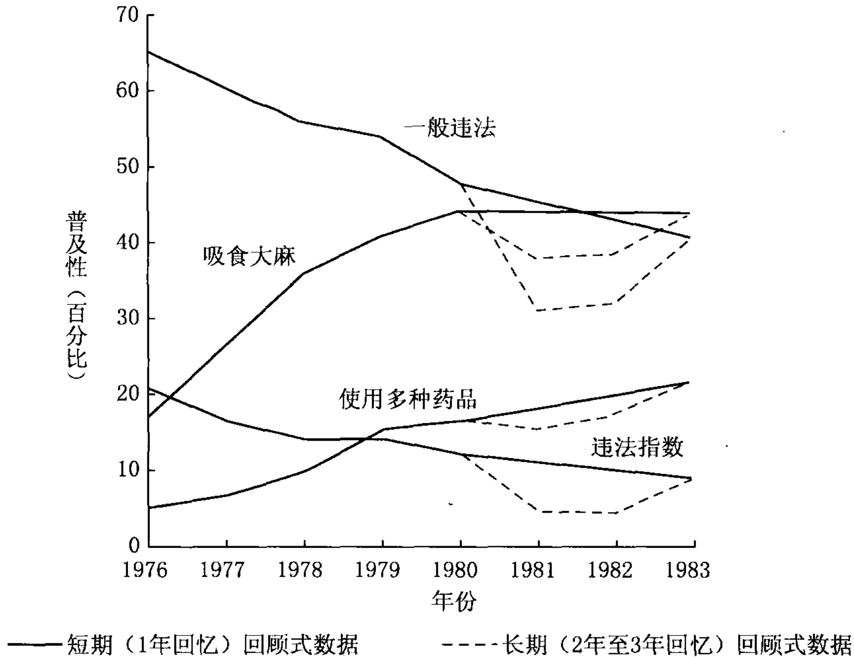


图 4.1　基于长期与短期回忆的时期趋势

1981 年至 1983 年的前瞻数据(基于 1980 年和 1983 年)显示了一般违法的稳定或下降趋势,但违规指数和使用大麻的长期回忆数据(1981 年和 1982 年)显示了上升趋势。对于使用多种药物,长期回忆数据与 1 年回忆的数据偏离不大,但对全部非法行为,有一个"北斗七星"的效果,长期回忆的趋势低于 1 年回忆的趋势。图 4.1 符合了人会随着时间而慢慢忘记过去的事件和态度的情形。记忆衰退的问题相对于其他问题(一般非滥药罪行),对某些违法行为(滥用硬性毒品)而言并不太严重。梅纳德和埃利奥特提出过其他的解释,但得出的结论还是记忆会随时间而淡去是最可信的解释。

比较前瞻(1 年回忆)和回顾(10 年回忆)的数据,相同的受访者被问及是否(如果有)曾经犯下某些严重罪行(强奸、抢劫、严重侵犯、结伙斗殴、入室盗窃、盗窃超过 50 元、偷车、销售硬性毒品)。简单来说,前瞻自我报告包括超过 95%的回顾自我报告罪行,但回顾自我报告包括前瞻自我报告的不到一半。这些结

果验证了记忆随时间而淡去的假设。梅纳德和埃利奥特检验了其他可能的原因,但是没有发现比忘了几年前发生的事情更合理的解释(Menard & Elliott, 1990a)。虽然这些结果并不排除使用回顾数据,但是总的来说,关于行为的长期回顾数据显然是不可靠的,可能会产生与前瞻数据相反的趋势或影响。

复制和回忆的问题可能还包括伸缩和逆伸缩现象:报告某事件在某段时期的前后再次发生。全国犯罪调查研究有伸缩现象(Lehnen & Skogan, 1981)且其受访者的第一个访问并不用以样本的受害率估计。其他研究尝试使用记忆包围技术(参考事件,如生日或假日)以减少伸缩和逆伸缩的现象。全国青年调查的数据也显示回顾与前瞻数据有伸缩和逆伸缩的现象(Menard & Elliott, 1990a)。

减少受访者在回忆事件时产生误差的方法包括:使用介绍材料、提出更详细的问题(而不是只问受访者一个是否曾经触犯法律的问题,需要提出一系列具体问题);提供清晰又容易记忆的参考时间(如假期、总统选举或重大新闻事件);试图将所问的事件与受访者生活中突出的事件相联系,例如向受访者展示日历并要求他们回答该期间内的重要生活事件(Converse & Presser, 1986; Fowler, 1998)。最后的方法就是"生活史日历"(Freedman et al., 1988),即研究人员向受访者展示一个日历(受访者要回答每个事件横跨的日期或有关的特点),并要求他们指出所要研究事件的发生日期。弗里德曼等人指出,这种做法耗费很大,而且编码会很繁琐,但较适合研究回顾事件回忆(例如,生小孩或搬家)或环境变化(例如,入学或婚姻状况)(Freedman et al., 1988)。相对于研究态度的数据,关于生命日历研究方法(LHC)的评论,塔里斯总结说:"公平地说,LHC 有时能增加对一些变量的回忆,但肯定不是所有的变量。"(Taris, 2000: 11)前瞻纵贯设计的数据通常比回顾设计更好、更可靠且更准确。

受访者回忆的问题主要在于微观社会访谈研究。使用前瞻追踪样本设计有助于减少这些问题,但不能完全消除它们。回顾设计可能存在效度的严重问题,因为随着时间的长度增加,受访者对报告的事件或行为的记忆越来越模糊(Weis, 1986)并可能漏报。基于某些目的,如认定测量态度随时间而改变,只用前瞻追踪样本设计已足够。最后,当使用重复截面设计来测量随时间而转变的

问题时要特别留心。即使抽样程序或进行调查问卷只有轻微的差异，也可能会导致复制时出现严重的问题，如马丁的例子。

纵贯研究的成本

沃尔和威廉姆斯认为，前瞻追踪样本研究每一轮的费用并不比相似数量的截面研究高（Wall & Williams，1970）。总共6轮的前瞻纵贯调查的费用可能不会超过6个相似人口或样本数量的截面研究。即便如此，仍必须考虑一个6轮前瞻研究的费用是否比6个独立的截面研究或是回顾追踪样本研究小组的6倍高。并非所有的研究都需要纵贯数据，有些研究可以用其他研究人员收集的纵贯数据进行二次分析。

对具有某些目的的研究，纵贯研究是唯一的方法。如果研究目的是衡量历史或发展的变化，纵贯设计是必不可少的，尤其是独立的年龄、时期及世代效应。如果衡量长时间的变化，那么前瞻追踪样本设计或总人口设计通常是最合适的，因为独立样本可能彼此不同，除非每一轮的抽样和数据收集的正式和非正式程序都可复制（Martin，1983）。此外，回忆问题可能会令回顾追踪样本设计的推论无效。如果衡量一段相对短的时间（周或数月）的变化，那么回顾设计可能也适用于事件或行为的研究，但不能用于态度或信念的研究。重复截面设计或循环追踪样本设计可能适用于由于反复访问而出现追踪样本条件习惯效应的问题。

如果研究目的是确定或估计因果关系的强度，那么纵贯研究就比截面研究好，尤其是当变量真正的因果关系和变化的时间序列未知的时候。理论上更糟的是，对因果关系和时间序列的猜测不足以验证真正的因果关系和时间序列。时间序列必须通过测试以确定是否存在假设的因果关系。最好的测验因果关系的方法是实验设计（Bickman & Rog，1998；Campbell & Stanley，1963；Rossi et al.，1999），而实验设计通常都是前瞻纵贯设计。

如果测量的变化不是重点，同时因果关系和时间序列已知，或研究不关注因果关系，那么，利用截面数据来分析可能已经足够了。但是，如果研究的问题

需要纵贯数据来分析，那么当然应该投入较多资金来得到正确的答案，而不是花较少的钱来得到可能错的或难以定论的答案。在这种情况下，纵贯研究的费用是否值得的问题与使用纵贯或截面研究无关。纵贯研究的成本是否由研究问题的重要性来决定呢？这就是正确地做研究或完全不做研究的选择。

第 5 章 | 纵贯分析

第 2 章提过纵贯研究的两个主要目的：描述变化的模式和分析因果关系。最后一章将更广泛地介绍分析方法以达到上述目的。因此，这章的重点将从纵贯数据的收集移到纵贯数据的分析，也就是数据收集好之后，我们要如何分析。很多地方对不同纵贯数据分析的方法都有全面的描述，这章将不会详尽地展示如何使用每个方法，而是着重介绍纵贯研究会处理哪些不同类型的研究问题以及其回答研究问题的不同方法。更详细地解释纵贯分析方法的研究可在不同方法的引用文献中找到。纵贯研究的介绍请参阅塔里斯(Taris, 2000)等人的著作。

纵贯研究的第一个目的是描述变化，重要的是要区分以下几点：(1)定性和定量的变化；(2)短期(几段时期，无论每段时期的实际长度)和长期(很多时期)的变化；(3)我们所关注的改变是一个或多个变量的改变还是两个或两个以上变量关系的改变；(4)我们是否有兴趣描述、预测或解释变化。关于第四点，描述改变通常涉及时间或年龄和一些变量的二元关系。预测变化可能涉及一个基于时间的简单的趋势预测，或可能涉及更复杂的多预测变量的模型。解释变量需要附加关于变量内因果关系的假设和理论，这就带来纵贯研究的第二个目的——纵贯研究的因果分析。

在因果分析中，我们还需要区别几点：(1)定性和定量的结果；(2)短期和长期的分析；(3)我们是否对某特定结果、两个或两个以上结果的彼此关系有兴趣(请注意，这里使用"结果"而不用"变化")；(4)我们感兴趣的是一个变量随时间而变化还是个案之间在某个时间点上的结果的不同。无论这个变化是个体内(例如当个案是人时)还是个案间的差异(如个体间的差异)，都是因果分析所关

注的。举例来说，我们更感兴趣的是国家之间总生育率不同的原因（因为它可能受到人均国民生产总值、识字水平和家庭计划公共资源的影响），还是某个或某些国家推行新的计划生育方案后的变化？同样，我们对青少年吸食大麻频率的差异（例如，他们一年内吸食的次数）更感兴趣，还是对个人吸食大麻的频率会随着年龄或者参与反毒品计划而变化更感兴趣呢？这些明显是不同的问题，对如何衡量和分析一定会有不同的影响。

在这里简要说明 3 个问题：实验与伪实验研究、发展研究、差异的因果分析。大多数的实验与伪实验研究更关注短期变化。首先测量最早/底线的情况，然后对某些对象加入干预措施（治疗组），某些对象不加（控制组），在估计干预措施产生某些效果后，再做第二次测量。研究关注的问题是治疗组变化的方向或强度是否与控制组不同。社会科学的实验与伪实验往往有一个很实质的重点。例如，学生在新课程中会不会比在标准课程中学到更多东西呢？失业青少年在接受求职技巧培训后，会不会比没有培训的同伴更容易找到工作呢？这些问题的重点通常是短期变化的分析。

虽然发展研究有时会被当做变化的短期模式，但是它更关注长期性的变化，特别是整个人生或人生重要时刻的改变。例如，从学前到高中毕业的定量技能的发展，或从青春期前到中年或老年参与非法的行为。研究兴趣不在于两个特定时期的变化，而是变化的长期模式：是否增加和从未减退，即使它在某时间点已保持稳定（如我们所期望的发展数量的技能）或先升后降（这是典型非法行为模式）。除了模式的描述，我们也希望了解模式本身（结果和时间之间的关系）或非法行为的数量技能最终的水平（最终的能力水平）或参与情况（例如，在 21 岁测量的结果的水平和数值）会不会因不同特点的人（例如，性别、种族、社会经济地位）而出现系统的差异呢？发展研究对个人研究特别有用，因为个人有比较明确的开始（出生、第一年上学、第一份工作）和结束（死亡、完成正规教育、退休）。另外，它也可能对宏观社会研究有用，如城市或国家，这些研究对象不一定有明确（或至少不能明确地观察到）的开始和结束，特别是研究兴趣所在的变量（生育率和死亡率、识字率和人均国民生产总值，与个人层面的例子相同）。然而，在个人和宏观社会层面中，我们有时对随时间而变化的模式不感兴趣，反

而更在意特定的年龄或时间的结果水平或数值。典型的例子是布劳和邓肯的研究(Blau & Duncan, 1966),即男性在某一时间的职位是由父亲的教育和职业、自身所受的教育、第一份工作的职位构成的函数。虽然这里也关注到变化(职业地位的变化:从父亲的工作到儿子的工作,或从儿子的第一个工作到目前的工作),但是最大的重点是研究获得的地位(Grusky, 2001),为什么有些人比其他人更容易升任更高的职位?这种分析称为"差异因果分析"(暗示个人之间的差异,并非个体内随时间而发生的变化),这与变化分析(因果或其他)不同。

很重要的一点是,虽然应用/实验和发展研究要求对相同个案进行反复测量(无论是个人或总人口),但是反复测量只是一种选择,并非差异因果分析的必然要求。原则上来说,可以(并且已经)使用时间序列截面模式,例如,前文提及的职位地位以及跨国生育率、家庭计划与发展的研究(Tolnay & Christenson, 1984)。无论是基于总人口、重复截面还是回忆追踪样本设计,纵贯数据仍然可用于差异因果分析,并提供优于纯粹截面(或时序截面)数据的数据。

纵贯与截面统计模型

假设我们选择一套理论上合适的因变量和自变量。进一步假设,基于理论与过往的研究证据,我们相信变量间的因果次序,并且可以适当地转换变量,把因果模型嵌入一般线性模型(例如,潜变量结构方程、多元回归、方差分析和协方差分析、logistic 回归或判别分析)来识别模型(Heise, 1975)。现在假设我们要计算直接因果关系的强度(如果我们使用路径分析或潜变量结构方程模型,我们也可以计算间接效应),为什么该研究需要纵贯数据,而不是截面数据?

勋伯格表示,在某些条件下,利用截面数据做应用动态模式能得到有效、无偏差的模型参数估计(Schoenberg, 1977)。该基本条件是"非遍历",也就是说,它依赖于该系统的初始状态。对于遍历系统,系统不依赖于该系统的初始状态,但会产生无论在哪一段时间都相同的系统,其基于截面数据计算的动态模型,其参数估计可能会有所偏差和无效。

遍历系统的例子是,外生变量的变化是随机的,其期望值在任何时间 t 和其

他任何时间 $(t-k)$ 上都相同。在数学上，$E(X_t)=X_0$，X_0是 X 的初始值。在非遍历系统的例子中，自变量的变化不是随机变量，是依靠外生变量之前的数值。换言之，在自回归过程中，随机变异、外生变量产生的数值和 X 值都不是常数。在数学上，$E(X_t)=\sum\phi_k X_{t-k}$，其中 $k=1, 2, \cdots, K$，X_{t-k}是X 过去的数值，ϕ_k是 X_{t-k} 的系数。截面数据是否适用于计算外生变量(X)和因变量(Y)的动态关系取决于哪个过程导致了 X 的变化，是来自随机变异还是自回归。相反，纵贯模式可用于遍历和非遍历进程。

第二个使用截面数据去估计纵贯模型参数的潜在问题在非尔鲍的生育率和识字水平研究中就已说明了(Firebaugh, 1980)。表 5.1 和图 5.1 取自非尔鲍的研究(1980:340—341)，说明了截面和纵贯相关性的方向可能是相反的，然而两个结论可能都是对的。从 1961 年至 1971 年，截面数据显示，生育率最高的地区是印度的旁遮普邦，该区识字率也是最高的。但是，在每个地区内，随着识字率的提升，生育下降了。在这个例子中，截面和纵贯数据产生了非常不同的关于生育率与识字水平关系的结论。如前文提到的，梅纳德和埃利奥特以及格林伯格的关于年龄与非法行为关系的研究也有相似的发现。如非尔鲍所言，确定这两个模式中哪个更合适或更重要是理论上的问题，而不是研究实验的问题。不过，这里要说明的是，截面数据并不经常用来模拟动态、纵贯关系。梅纳德和埃利奥特的截面和纵贯数据分析再次加强了这个论点。另外，戴维斯和毕扣斯的动态模型模拟研究也说明了截面分析未能正确地推算预设人口参数，在一定的抽样误差内，纵贯分析的估计是不错的(Davis & Pickles, 1985)。

表 5.1 生育率与识字水平的截面与纵贯相关性：印度旁遮普邦(1961—1971 年)

时间系列相关(同区域不同年份)		截面相关(同年份不同区域)	
区域	相关	年份	相关
阿姆利则	−0.9	1961	0.5
珀丁达	−0.5	1962	0.6
菲罗兹布尔	−0.9	1963	0.4
古尔达斯布尔	−0.9	1964	0.6

续表

时间系列相关（同区域不同年份）		截面相关（同年份不同区域）	
区域	相关	年份	相关
霍希亚布尔	－0.8	1965	0.1
朱伦杜尔	－0.7	1966	0.2
格布尔特拉	－0.4	1967	0.3
卢迪亚纳	－0.9	1968	0.6
帕蒂亚拉	－0.3	1969	0.5
罗帕尔	－0.4	1970	0.6
森格鲁尔	－0.1	1971	0.7

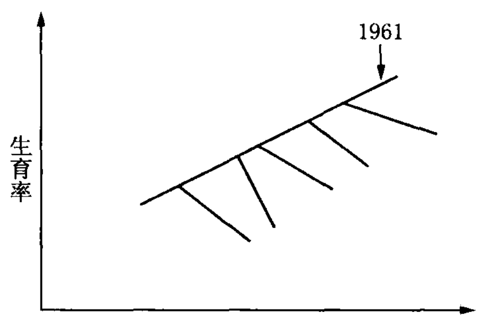


图 5.1　生育率与识字水平之间的关系图

纵贯因果模型的种类

因果模型有 4 个“纯粹”的类型：(1)$X \rightarrow Y$，因变量的数值可用自变量数值函数来表达；(2)$\Delta X \rightarrow Y$，ΔX 代表 X 值的变化，因变量的数值可用自变量变化的函数来表达；(3)$X \rightarrow \Delta Y$，ΔY 代表 Y 值的变化，因变量的变化可用自变量数值的函数来表达；(4)$\Delta X \rightarrow \Delta Y$，因变量的变化可用自变量变化的函数来表达。混合模型的自变量包括变量的水平和变化率（例如，人口密度和经济发展的人口增长率）。

在截面数据的因果分析中，我们通常假设一个测试模型 D：X 的变化产生

(导致、引起)Y 的变化。但更常见的是，我们会测试截面和纵贯分析中的模型 A。准确地说，模型 A 表示，一个变量(因变量)的水平或数值取决于一个或多个自变量的水平或数值。换言之，X 的差异导致了 Y 的差异，例如，薪金差距导致不同程度的工作满足感，或识字水平导致不同的生育率。这暗示了因变量的变化同样是自变量变化的函数，例如，薪金变化产生工作满足感的变化，但要检验的是模型真实方程中所涉及变量的数值，而不是数值的改变。

另外，一个变量的变化可能导致另一个变量的一系列变化，但这两个变量的数值在很大程度上不相关。这说明了 X 和 Y 的初始值是不相关的，但 X 的变化导致了 Y 的变化。例如，营养水平(以人均卡路里供应计算)最初与国家家庭计划推行力度(高和低营养水平的国家都会积极或不太积极地推行家庭计划)无关，但增加家庭计划的力度可能会导致人均卡路里供应的上升，因为个人有意或无意地根据他们自己国家或地区的容量，更有效地调节自己的生育能力。积极的家庭计划未必与高营养水平相关，但提高家庭计划的措施会与改善营养水平有关。这个过程虽然违反直觉，但仍有可能。如果存在这样一个过程，那么模型 D 是恰当，但模型 A 就不可以。另一方面，适合模型 A 的任何情况都可能适合模型 D。如果一个变量会影响另一个，那么第一个变量转变了，第二个变量必然会变(即，如果模型 A 正确，那么第二个变量的变化会取决于第一个变量的变化)。从分析方面来看，使用一阶差或无条件变化模型可能产生模型 D(Finkel，1995；Liker et al.，1985)。

纯模型 B 的例子是，压力水平(因变量)随着收入而改变(不论初始收入水平如何，大量加薪会产生低压力，大量减薪会产生高压力)。模型 C 表明，变化率取决于某些差异，可能是稳定的差异。例如，薪金上涨可能基于某些固定的特点，如种族(在有种族偏见的组织内)或教育程度，最高学历或完成的学位(在精英组织内)。模型 C 可以是其中一种差异方程模型。例如，理查德森使用双差分方程解释国家之间的军备竞赛(Richardson，1960)：

$$dx/dt = kY - aX + g$$

$$dY/dt = jX - bY + h$$

其中，X 和 Y 代表两个国家的军备水平，j 和 k 代表非负的“防务”系数（鉴于别国的军备水平而感到需要武器），a 和 b 代表非负的“疲劳”系数（国家经济衰退或已经装备部分武器而不愿加强军备），g 和 h 是常数，国家彼此之间的正或负的“不满”因素可解释为敌意（如果是正向）或友善（如果是负向）。方程左边的 dX/dt 和 dY/dt 代表随时间而变化的比率，右边代表 X 和 Y 水平的变化。

穆尔丁和贝雷尔森在他们的回归分析中应用模型C解释较不发达国家的总出生率变化（Mauldin & Berelson，1978）。总出生率的变化为因变量，家庭计划方案和社会经济发展指标为自变量。徐和博格进行了类似的分析，不过他们用滞后内生变量的模型A（Tsui & Bogue，1978）。马库斯解释道，以改变值（Y_2-Y_1）为因变量与用滞后内生变量一样，只是假定模型中内生变量的滞后系数为1。对于滞后内生变量模型，因变量 Y 和自变量 X 满足 $Y_2=a+bX+cY_1$ 或 $(Y_2-cY_1)=a+bX$。如果 $c=1$，则方程为标准回归模型，改变值 (Y_2-Y_1) 为因变量（Markus，1979）。鉴于变化数量和滞后内生性变量的方法相似，穆尔丁和贝雷尔森及徐和博格得出了几乎相同的实质性结论，连解释因变量的方差都几乎相同。马库斯认为，对上述模型而言，限制系数 c 等于1并没有任何好处，他建议因变量最好是滞后内生变量，而不是改变值。

在所有模型中，特别是模型A和模型D，测量的时间是一个问题。当 $X\to\Delta Y$ 时（模型C），假定 X 的测量应该早于 Y 变化的一段时期；当 $\Delta X\to Y$ 时（模型B），测量 X 转变的时间要先于 Y 的衡量。不过，对于 $\Delta X\to\Delta Y$，个案可能作出改变以同时测量 X 和 Y 的变化（实时效应），测量 X 变化的结束时间应早于 Y 变化的开始（滞后效应）或在前两种可能性之间，测量 X 在某时期的变化应早于 Y 变化的开始，但在 Y 开始变化后就得结束（部分滞后效应）。同样，$X\to Y$ 时（模型A），我们需要考虑 X 是否该在 Y 衡量（滞后效应）前衡量，或同时衡量（实时效应）。

量度变化

量度或描述最基本的变化会涉及两个变量的二元关系，一个是量度时间序列或年龄，另一个是量度一些兴趣所在的结果。这个看似简单的任务有时却很

困难，特别是在选择最适当变化的测量的时候。首先，需要区别定性和定量的变化。定性变化的测量很直接：或存在变量数值上的改变（即质性状态），或没有变化。例如，一个人从没有违法行为（不参与非法活动）到违法行为（参与某些犯罪活动），或者一直保持不违法。一个人从蓝领工人升到白领工人，或仍是蓝领工人。对于变量的每个单独的类别，形式是相同的：人改变或不改变。因此，测量纯粹的质性变化可能只需要简单的二分法："有"或"没有"。如果类别是定序的且组类很少，那么用二分变量测量变化就可以了，但是如果是包含很多组类的定序变量，那么测量变量就需要一个更详细的定量测量。我们更想知道个案"如何"改变（例如，针对"宗教这个名义量表"，我们可能想知道一个人是否从基督教转为天主教或犹太教），但原则上，这只意味着根据变化的不同二分可能性。不管有没有改变，它仍是二元测量。对于连续量表的变化，通常应用两种测量方法。第一种是变量前后分数的差距 $(X_2 - X_1)$，下标指的是时间。这称为差异、变化分数、原变化或原增益。第二种方法是研究改变是残差增益。为了计算残差增益，变量 Y_2 首先对 Y_1 进行回归（使用线性回归），以获得 Y_2 的预测或期望值。Y 的期望值 $E(Y_2)$ 是基于 Y_1 的数值和两个参数的值 a（截距）和 b（Y_2 和 Y_1 之间关系的最佳拟合直线斜率）：$E(Y_2) = a + bY_1$。残差增益数值是 Y_2 的实际值与其期望或预测值之差：残差增益 $(Y) = Y_2 - E(Y_2) = Y_2 - a - bY_1$。

对于比例量表的变化，第三种常用方法就是 Z 变化的百分比：(Z) 的变化百分比 $= 100\%(Z_2 - Z_1)/Z_1$。这个方法并不适用于其他量表，因为任何没有非武断零点的量表都存在无限数量的变化，但与百分比变化同样有效。为了说明这点，温度便是一个好例子。在这个区间量表中，室温到水的沸点的百分比变化是 $100\%(212 - 70)/70 = 203\%$（华氏量表），但摄氏量表则为 $100\%(100 - 21)/21 = 376\%$。另外，当 $Z_1 = 0$ 时，变化百分比就计算不了，因为不能除以 0。变化百分比的变异作为变化的衡量，包括基于比率（而不是百分比）和变化的复合比率，最常见的例子是储蓄户口的复合利息。衡量的变化可能基于多于一个的时间单位。计算方法很简单，就是改变除以时期的次数（或其他时间单位，不一定与测量时间长度相同），如此便得出变化率（各时间单位的原增益、残差增益或百分比变化，如犯罪率或人均收入每年平均百分比的改变）。

有3种描述变化的常见形式:数值、图形或数学(包括统计)。变化的数值描述包括测量的数值变化,例如,人均国民生产总值每年的百分比变化。简单测量的变化(如短期实验与伪实验研究)存在更多相似点,与发展研究的相似点较少。图形描述一般将变量在不同时期的数值对时间作图,水平轴是时间,竖直轴代表变量。图4.1就是变化图形描述的例子,表示在特定时间内,趋势是向上、向下还是稳定,以及在不同的时间,趋势是否变化且如何变化。变化的数学和统计模型描述主要用方程表达。量性变化模型的重点在于,一个变量随着时间而改变多少,而质性变化模型的重点在于,个案随时间而改变多少。

变化测量的问题

以上提到的任何变化测量的方法都可用于总人口设计、追踪样本设计和循环追踪样本设计的个人或团体数据。在重复截面设计中,我们不太可能去测量个别变化,但可衡量界定清晰的组群的变化,只要是从组群层面上,这些个案在不同的截面中就是可比的。对每组的概率抽取足够个案,虽然取样和数据收集程序要严格复制,但任何对最初的抽样或行政程序的偏离都可能严重影响数据的可比性,同时令重复截面数据无法进行纵贯分析(Martin, 1983)。图3.1的纵贯设计描述了可以用来衡量总样本或人口的变化,但重复截面设计必须确保不同截面的抽样和管理程序相同。

每当我们试图衡量变化时,我们需要考虑从一个时间到另一个时间的明显差异是否真的代表变化且是不是不可靠的测量。之前的举例就说明了这个问题。雷德菲尔德和刘易斯对墨西哥村庄迪坡斯特兰的研究就有不同的结论,是因为不可靠的结果(一方或双方观察员有偏见,并做了不准确的解说),还是因为从雷德菲尔德到刘易斯这段时间里,村子发生了重大的变化?全国犯罪调查的访问次数与受害事件数量呈负相关是否因为该研究调查数据不可靠,或是真正的下滑(可能由于“治疗”效应,访问调查使受访者想起了受害的经验,从而提升警觉而避免再次受害),或存在抽样问题(高受害者可能因为经常搬家而流失),或是受害的真正时期趋势呢(Cantor, 1989)?样本特色在1976年一直很稳定,同时,全国犯罪调查中大部分罪行的估计受害率在1976年后下降了

(Bureau of Justice Statistics, 1992; Rennison, 2000)。因此,重复访问的受害率下降可能是因为真正的时期趋势,但这又不太可能解释所有观测到的变化(Cantor, 1989)。有趣的是,受害趋势可能比受害水平更准确,因为自1976年以来,全国犯罪调查的样本内的误差大概都有相同的分布。

回想一下再测信度(不是内部一致性可靠测量),这是测量本身的变化,但变化实际上没有发生(Zeller & Carmines, 1980)。区分不可靠的测量和实际变化的方法是复制和使用多个独立可靠的量表。在某些情况下,其他证据能清楚地反映哪一个解释比其他的好。例如,特定年龄分析显示,年龄组的非法行为普及率的变化不多,这些变化可能与以前的访问次数无关,这表示了真实变化,而不是不可靠测量。如果各年龄组的变化不一,同时这些变化似乎与之前的访问次数相联系,那么这很可能是不可靠测量。

描述变量之间关系的变化

直至现在,重点一直在变量数值如何随着时间而变化,这通常是变化描述的主要关注点。但是,研究变量间的关系也可能随着时间而变化。佩恩等人举了两个例子,一个例子使用 logistic 回归观察工作资格和失业关系随时间而发生变化,第二个例子使用对数线性分析来检验社会阶级和政党间的关系如何随着时间而改变(Payne et al., 1994)。在第一个例子中,他们发现,个人就业资格的不足随着时间而变化,更具体而言,该变化在高失业率的早期比低失业率的后期更大。第二个例子与前面提到的豪特等人的研究相似,他们发现在英国选举中,社会阶级和政党的关联随着时间而减弱。这些例子都涉及因果关系或至少预测模型,但请注意,重点在于因变量或结果变量与其预测变量之间关系的变化,等同于甚至超出预测变量预测结果变量的能力研究的重要性。此外,在一般关系变化的分析中,有些人认为该变化是其他变量的历史或发展变化。

类似的分析可以采用结构方程建模,例如,使用堆栈模式在不同时期测试结构参数的等同值。尤其重要的是测试因子不变性(例如,Kaplan, 2000; Patterson, 1995),即同样的指标是否在不同时间具有相同的概念?这个因子不变性的问题是,测量在不同时间是否恒定?如前所述,即使在不同时间使用相

同的测量，但这并不表示我们每个时期所测量的事情都一样。帕特森举了一个反社会特征的例子，个人行为表现指标随年龄增长而改变，并说明不同的偏差行为如何改变，如滥用药品和与警察接触的重要性都提升了。如帕特森所说，尤其是生命过程的研究，不同的变量可以衡量在不同的发展阶段的相同事物(Patterson, 1995)。

决定性与概率模型

数学和统计模型分为决定性模型与概率性模型。在描述变化的决定模型中，所有个案既有特点的改变需要完全一样。在决定的因果模型中，知道预测变量的相关有限集合的数值或某特定个案是否改变或改变多少，能令我们知道(原则上)预测值或因变量的确定值。它是否会改变？如何改变(增加或减少)？同时，个案会改变多少？实际上，社会科学的决定模型有可能出现偏差，如测量错误。在社会科学中，定量变量的变化决定性模型包括函数方程(Kim & Roush, 1980: 101—104)、差异方程(Huckfeldt et al., 1982; Kim & Roush, 1980:第5章)和微分方程模型(Blalock, 1969:88—91; Kim & Roush, 1980:第6章)。这些模型所表达的变量数值的变化可被描述为时间的函数。变量变化的描述应包括该变量和数学公式的时间项；变化的解释涉及方程中的其他变量；变化的预测可能涉及时间以外的预测变量。

前文提及的理查德森的军备模型(Richardson, 1960)便是一个决定因果模型，它试图解释军备水平的变化。另一个变化决定描述模型的例子是内部—影响扩散模型(Mahajan & Peterson, 1985)。创新的扩散简单模型，如内部—影响扩散模型，通常把在一个特定时间内采用创新的累积数目表达成时间的函数，可表达为微分方程(Hamblin et al., 1973; Mahajan & Peterson, 1985)。一个可能描述这个过程的方程是 $dX/dt = ct^n$，其中 X 是采用创新的累计数目，dX/dt 是 X 的变化率，t 是一些适当单位的测量时间，而 n 和 c 是需要估计的常数参数。如果我们把方程整合一下，可写成 $X = ct^{n+1}/(n+1)$。最简单的形式是，当 $n=0$ 时，方程变为 $X=ct$，X 便为时间的线性函数，常数 c 可用普通最小

二乘回归技术估计。当个案数量相对较大(例如,超过 20)且测量期间相对较小时,这种做法会随着 t 多项式函数的变异而有效地描述变化。

在变化概率描述模型中,我们期望所有个案的变化不都一样,即使它们具有相同的特点,但一定比例的个案会以某种方式改变。在概率因果模型中,如果我们知道预测变量的一些相关、有限集合的数值,或它们如何应对总人口、样本或足够大的组群的变化(男性和女性、较发达和不发达国家),我们就可以较准确地预测具有某个结果的个案的比率或百分比或它们是否会改变、以某种方式改变的比例或百分比、改变的平均数量(平均数、中位数、形态)。我们无法预测个别个案的结果,或结果是否、如何改变,或改变多少。因为概率模型的基本假设是,对个别个案行为的影响是一个概率的过程。具有某些特征的个案或多或少比具有其他特点的个案更会出现某特定方式的改变,但组群中个案的改变模式不必一致。如前所述,穆尔丁和贝雷尔森及徐和博格的生育率模型便是概率模型。

观察概率模型的一个方法就是认定个人的行为表现受到不同的影响(对某些人或组群在某个方面有强烈的影响,但对其他人较弱),但同一时间又有自由去选择不同的行为模式,甚至抗衡强烈的影响。有些人会选择抵制对他们的行为有可衡量的影响,即使这些影响是强大的,但能抵挡强大影响的个案比例比弱小影响的少(这意味着强大的影响力有较小的预测误差)。

社会科学中的决定模型比较罕见,通常用于模拟定量变量。概率模型,尤其是统计模型,在社会科学领域中是很常见的,它广泛运用于定性和定量变量。本章后半部分将集中讨论利用统计模型分析社会科学的纵贯数据。选择合适的模型来分析一个具体的研究问题,取决于现象问题的假设能多准确地被衡量以及数据收集的设计和实施的强度,但选择纵贯数据的因果分析的好方法是考虑个案数目(n)和时期数目(t)。表 5.2 就这两个要点提出了一个粗略的分类。请注意表 5.2 的明显的“灰色地带”,就是 20 至 100 的个案(取决于分析方法)和 10 至 20 个时段的情况。在这个灰色地带,选择适当的方法是艺术而非科学,并且最好能根据经验和熟悉的方法,而不是任何规则。

表 5.2 分析模型和数据结构

少量个案 ($n < 20$) 很多时期 ($t > 20$)	大量个案 ($n > 100$) 很多时期 ($t > 10$)
ARIMA 模型：共变量、转换功能模型、干扰时间序列模型 自动回归(AR)时间序列模型 滞后内生变量(LEV)模型 分类数据(最佳量表)的多元动态分析	连续时间事件史分析：高斯比例风险和参数风险模型 多层次增长曲线模型
少量个案 ($n < 20$) 很少时期 ($t < 10$)	大量个案 ($n > 100$) 很少时期 ($t < 10$)
集合截面/时间序列分析	线性追踪样本分析条件变化模型(滞后内生变量) 线性追踪样本分析非条件变化模型(变化分数) 潜在增长曲线分析 分离时间事件史分析 多层次增长曲线模型

集合截面和时序数据

集合截面和时间序列数据需要了解两个不同的方式：数据结构和分析纵贯数据方法。作为一个数据结构，如图 5.2，时期“堆积”起来，就好像它们是同截面的某些部分。表中的每个 X 代表了某变量在某特定个案中于某特定时间的一个观察值，X 下标的符号 $n = 1, 2, \cdots, N$ 个个案，$t = 1, 2, \cdots, t$ 时期，$k = 1, 2, \cdots, K$ 个变量。在图 5.2 中，列代表个案，行代表变量，除了那些重复 T 次的个案(行)。因为只有一个因变量 Y，所以没有必要设第三个下标符号，但原则上可以有多于一个的因变量。同时，如重复截面设计，个案的每个时期的衡量有可能不同，在此情况下，便只是集合截面，而不是集合截面时间序列的设计。重复截面数据结构对分析某层次(国家、城市等)样本的集合历史转变很有用，同时可以用熟悉的分析方法，如普通最小二乘法或 logistic 回归，但如前所述，

它不允许非集合个案的改变的测量和分析。当反复衡量同一个案时，如纵贯追踪样本设计，集合截面时间序列（或 TSCS）数据结构提供了更好的统计效能和更可靠的估计，但与任何分析一样，其缺点在于参数估计可能被不同时期个案内或同一时间测量个案之间的真实值或/和误差的相关所混淆。TSCS 数据结构是较成熟的纵贯数据分析的标准格式，包括潜在变量成长曲线模型、多层次增长曲线模型和事件史分析。这些技术（下文将会讨论）需要大量的个案数量和时期，才能得出可靠的模型参数估计。表 5.2 左下方代表最坏的情况，只有少数个案和少量时期。在这种情况下，我们只可能做案例研究描述，即使简单的统计描述也不太合理。相反，如果有大量的个案和时期，就可能像图 5.2 那样建构数据，并使用较简单的普通最小二乘法回归或类似技术来分析数据。塞尔介

		变量 1：X_1	变量 2：X_2	…	变量 K：X_K	因变量：Y
时间 1	个案 1	X_{111}	X_{112}		X_{11K}	Y_{11}
	个案 2	X_{211}	X_{212}		X_{21K}	Y_{21}
	⋮					
	个案 N	X_{N11}	X_{N12}		X_{N1K}	Y_{N1}
时间 2	个案 1	X_{121}	X_{122}		X_{12K}	Y_{12}
	个案 2	X_{221}	X_{222}		X_{22K}	Y_{22}
	⋮					
	个案 N	X_{N21}	X_{N22}		X_{N2K}	Y_{N2}
	⋮					
时间 T	个案 1	X_{1T1}	X_{1T2}		X_{1TK}	Y_{1T}
	个案 2	X_{2T1}	X_{2T2}		X_{2TK}	Y_{2T}
	⋮					
	个案 N	X_{NT1}	X_{NT2}		X_{NTK}	Y_{NT}

图 5.2 集合截面/时间序列的数据结构

绍了几种集合 TSCS 数据模式，其中最简单是常数系数模型（Sayrs，1989）。即使是在不同时期重复测量相同的个案，仍假设预测变量的测量和结果是独立的。但是，同样的个案在不同时间的测量不太可能完全无关。另一种模型就是最小二乘虚拟变量（LSDV）模型，它仍可使用最小二乘法回归来估计，同时，LSDV 模型的计算可以假设相关基于时间而非个案。但是，若假定线性关系同时基于时期和个案，模型便无法计算，因为这个结果将是完美的共线性。即使是 LSDV 模型，要准确地估计模型，个案和时间的数量仍必须超过表 5.2 左下方的最大值。

塞尔还介绍了更复杂的模型，包括广义最小二乘、随机系数和结构方程模型，但越复杂的模型，就需要越多的个案或时期来估计。成熟的纵贯分析技术如下：如贝克等人建议，政治学的 TSCS 模式一般涉及 10 宗至 100 宗个案，跨时 20 年至 50 年（Beck et al.，1998；Beck & Katz，1995）；若是二元结果的数据，他们建议使用事件史分析（时间是虚拟变量的比率模型，因此会包括 LSDV 的重要元素）；若是分类和连续结果，可以建构随机系数模型来计算个案内分层（即相同个案在不同的时间的多元观察）和观测值的时间依赖性（后者的预测变量加入时间部分函数，这可能与模型中的其他预测变量互动）；凯斯勒和格林伯格以及塞尔建议，较简单的 TSCS 数据的方法对少量个案和时期的分析会很有用（Kessler & Greeberg，1981；Sayrs，1989），如表 5.2 左下方的情况，但是，当个案和时间都足够大的时候，最好运用其他方法。

时间序列分析

表 5.2 的左上方是多时期少个案。在这种情况下，我们通常会研究一个或数个个案的改变模式（可能用非统计及非正式的个案比较），其重点是归纳跨时段的情况而不是分析个案。技术上来讲，当同一个案的同一变量具有两个或两个以上时期的数据时，我们就有一个时间序列。但是“时间序列分析”一词通常指分析具有长时间序列的单一个案的一系列分析方法。不同时序分析的类型包括自回归移动平均（ARIMA）模型（Box et al.，1994；Wei，1990；Yaffee & McGee，2000）、时间序列回归（TSR）模型（其他名称有“简单自回归”或“计量经

济时间序列模型”)(Ostrom, 1990; Yaffee & McGee, 2000)、滞后内生变量(LEV)模型(Sanders & Ward, 1994)、光谱分析。时间序列变异分析较之社会科学数据分析的其他几种方法较少使用(Jenkins & Watts, 1968; Wei, 1990)。此外,对于定性因变量,比莱韦德等人认为,可使用最佳量度技术去做分类数据的多元动态分析(Bijleveld et al., 1998:132—148),但这种技术比较少用。在“中断”时间序列分析(Cook & Campbell, 1979; Wei, 1990; Yaffee & McGee, 2000)的模型中,至少包括一个二分预测变量,这个变量的变化可能会影响因变量。

ARIMA 时间序列分析在社会科学中的应用越来越广泛,特别是在博克斯和詹金斯的著作(Box & Jenkins, 1970)出现之后。ARIMA 时序分析试图以 4 个过程的组合来描述长系列的时间次序。白噪音过程是一系列的随机冲击或改变,这是概率组件,出现在所有随机时间序列模型中。自回归(AR)过程是其中一个变量的当前值依靠其之前特定时间或区间的数值。移动平均(MA)过程是指白噪音过程的过往值继续影响模型变量在有限、指定时间或区间的当前值。综合(I)过程是指模型变量随着时间具有可检测的趋势或漂移,但系列没有趋势或漂移,这是因为它是变量值减去该变量之后的数值。减去或差异的目的是为了得到固定的白噪音时间序列,其中白噪音过程的数值的平均数为 0(即系列中随机组件在某一时间的值与同系列相隔某指定时间的值不相关)。一个时间序列分析可以结合一两个或全部 3 个过程以及白噪音过程,以得到一个固定的时间序列,并描述变量如何随时间而变化。此外,我们可以在 ARIMA 模型中结合连续或分类预测变量(后者如中断时序分析)(Sanders & Ward, 1994; Wei, 1990; Yaffee & McGee, 2000)。

TSR 模型看似一个最小二乘法回归模型,除了它是指一个个案而不是很多,而且是很多时期而不是一个时期,因为模型只对一个个案的因变量与预测变量做多次测量,但是,最小二乘法回归假设误差不相关通常都是错的。德宾-沃森(Durbin-Watson)统计用于测试自相关的错误,如果出现自相关错误,该模型就需要使用估计广义最小二乘(EGLS)或最大似然(ML)估计技术来重新估算。在模型中加入自回归组件,就会与 ARIMA 模型中的 AR 相同。基于使用

不同的估计方法(ML 不是 EGLS 做时序回归,它能对 ARIMA 模型做线性转换功能,但不做预白噪音化,详情见 Yaffee & McGee, 2000),ARIMA 模型和同样模型规格的 TSR 的结果可能相同。它也可以指定 TSR 模型,包括一个内生滞后的变量,然而需要进行不同的自相关误差测试(Ostrom, 1990:65—67)和不同的估计程序。TSR 需要较少时期,更重要的是,当预测变量多时,TSR 模型会比 ARIMA 模型容易(Ostrom, 1990; Yaffee & McGee, 2000)。许多人认为,ARIMA 模型(互相关函数决定因果次序)比 TSR 模型(因果次序是预先指定的)更受实验支配,但可以用格兰杰因果分析(见第 2 章)来决定 TSR 模型的因果秩序。

LEV 时间序列模型使用最小二乘法回归技术来预测结果 Y 的数值,这基于预测变量 X_1, X_2, …, X_K 的值以及 Y 之前的一个或多个值。在大多数情况下,模型会包含 Y 滞后一期的数值。桑德斯和沃德曾指出 LEV 的两个优势:第一,它通常避免无滞后内生变量的最小二乘法回归产生一系列相关误差的问题;第二,Y 的滞后值可以结合所有不可测量变量对 Y 当前值之前的影响(Sanders & Ward, 1994:203)。但是,误差可能会出现系列性相关或自相关(即在一个时间点的误差与该误差在其他时间点相关),而且预测变量对结果随着时间的影响可能不稳定。

TSR 时间序列测试并修正了序列相关错误,从而改良了 LEV 模型。ARIMA 模型透过差分不同时间序列和特定时期的随机冲击(移动平均或 MA 组件),不但能解释序列相关错误,而且说明了趋势或漂移(综合或 I 组件)。那么,为什么考虑 LEV 模型或 TSR 模型呢? 因为复杂模型要求更多的时期以得到更可靠的模型估计。ARIMA 模型,如前所述,时期需要可能多达 250 期,而 LEV 模型的估计,原则上需要的时期少得多。实践上来看,ARIMA 模型最常用于纯粹的描述或非常简单的改变模式(一个或两个预测变量)。TSR 和 LEV 模型往往用于较多变量和较少的时期。所有这些方法通常用于模拟定量因变量。原则上来说,这 3 个模型适用于分析质性因变量,但实际上,LEV 模型会更容易些(只用 ML logistic 回归或类似技术代替最小二乘法线性回归)。

时间序列分析长期以来被用于经济分析和预测以及自杀研究(Vigderhous,

1977)。桑德斯和沃德比较了最小二乘法回归、ARIMA、TSR(他们写为 AR),并用 LEV 模型来分析经济条件、消费者信心和福克兰群岛战争(Falklands War)对英国投票模式的影响。最小二乘法模型不理想,因为误差有高度自相关。更重要的是,其他 3 个模式会导致不同实质的结论。这 3 个模型有显著的自回归效应(依赖于前值或惯性),并且消费者信心对投票偏好有显著影响。只有 LEV 模型显示出经济条件对选民偏好有显著影响,并且,只有 TSR 模式无法找到福克兰群岛战争的任何长期影响。如何从这 3 种模型中进行选择?桑德斯和沃德对此总结道:"可惜没有一个容易或普遍的答案。"(Sanders & Ward, 1994:218)由于认识论的理由(LEV 过程似乎与个别选民的决策过程更接近),在该研究中,他们倾向于 LEV 模型,但他们补充说,这并不意味着 LEV 模型是最好的时间序列分析模型。

多个案短时间序列方法

表 5.2 的右下角可能是社会科学纵贯分析中最常见的情况,即大样本或个案数量多但时期少,这与之前的问题相反。对于量化数据越来越多地包括质性数据,可应用结构方程模型(SEM)做差异因果分析、线性追踪样本分析中的变化以及潜变量增长曲线模型的描述和解释。阶段性模型着重于离散、质性的状态以及包括马尔可夫链(Markov Chains)和对数线性模型之间的转换概率。实际上,所有模型主要用于短时间序列的分析。事件史分析和多层增长曲线模型(同样在表 5.2 的右下角)比较灵活,稍后在表 5.2 右上角的分析中会讨论到。

线性追踪样本分析

线性追踪样本分析(Finkel, 1995; Kessler & Greenberg, 1981; Markus, 1979)在第 2 章已描述过(测量时序、因果次序、线性追踪样本分析)。它可用两个时期,但很少用于超过 5 个时期的数据。线性追踪样本分析原则上可用来分析任何前文提及的基本模型(A、B、C、D),但最常用于模型 A——条件变化模型或模型 D——无条件变化模型(Finkel, 1995)。条件变化模型类似前节的

LEV 的时间序列模型，除了 Y 的前一个值会作为预测变量。在无条件变化模型中，Y 的变化 $(Y_t - Y_{t-1})$ 作为因变量，X_1，X_2，…，X_k 的变化作为预测变量，即 $(X_{1,t} - X_{1,t-1})$，$(X_{2,t} - X_{2,t-1})$，…，$(X_{k,t} - X_{k,t-1})$ 位于方程的右侧。线性小组模型可用来分析实验或伪实验数据，但至少有一个预测变量存在或缺少治疗或干预的情况，其他预测变量是共变的（其他预测变量影响治疗或干预结果），分析方法最好用最小二乘法线性回归、方差分析或协方差分析。这种类型研究最常见的模式是，Y 和它的预测变量都有两个测量周期，治疗前后各一次，Y 改变的程度和方向比 Y 在治疗后的实际价值更重要。另外，该模型可以用来分析非实验数据，最常用于结构方程模型（Bollen，1989；Hayduk，1987；Kaplan，2000），通常多于两周期。比起 Y 值的改变，该模型往往更注重 Y 值在预测变量不同值时的差异。

变值对滞后内生变量模型

社会和行为科学对改变值 $(Y_t - Y_{t-1})$ 作为衡量少量时间序列追踪样本数据变化的合适性一直没有达成共识。这些辩论主要是关于短期个体内变化的分析。对于较长时间序列的数据，当重点从短期变化的模式转到长期变化的模式时，该问题在很大程度上就不存在了（Bijleveld et al.，1998：39；Raudenbush & Bryk，2002：166—167）。克龙巴赫和菲比反对使用改变值（Cronbach & Furby，1970），因为改变值与测量随机误差会出现系统化的相关，通常比由它们计算出来的变量值（如 X_1 和 X_2）更不可靠，同时这不可靠的变化可能导致错误的结论或推论。他们还反对使用残差获得分数作为变量变化，残差获得分数只能基于最初的值分辨个案，这是替代变值的更合适的方法。普雷维斯同意上述观点并发现，测量误差与残差获得分数对改变值有同样严重的问题（Plewis，1985）。这些学者和其他作者都认为这是包括滞后内生变量 Y_{t-1} 的条件变化模型（Finkel，1995；Kessler & Greenberg，1981）。

另外有些学者赞成使用改变值，他们至少在某些情况下（Allison，1990；Liker et al.，1985；Rogosa，1995；Stoolmiller & Bank，1995）在无条件变化模型中用过改变值，其前提是，他们的研究兴趣在于个体内的变化而不是差异的

因果分析，时间序列数量较少（最典型的只有两三周期）以及满足某些其他假设。赖克表示，无条件变化模型在以下几个方面可能更优于截面方程和条件变化模型（Liker et al.，1985）：(1)回归参数从一个时期到另一个时期都保持不变；(2)不能测量变量影响因变量，但其不随时间变化；(3)变量测量有自相关的误差，这将影响因变量且不同时间有不同的变化；(4)追踪样本数据对预测变量随着时间变化的测量比特定时间的预测变量的水平或数值更可靠，因为个体间差异可能比最初的个体间差异大（Rogosa，1995；Stoolmiller & Bank，1995）。无条件变化模型比滞后内生变量模型好，这种情况是有局限的，大多数观察研究都不可能符合，也难以符合实验或伪实验研究（Cronbach & Furby，1970；Finkel，1995）。司徒米勒和班克认为，当个体间差异变化很小时，条件变化模型可能更有用（Stoolmiller & Bank，1995）。

艾利森也认同在伪实验不等控制组设计中的个体内变化研究可以使用无条件变化模型（Allison，1990）。假设任何治疗都没有稳定组群差异（即个案的不同"类型"），同时治疗和前测分数 Y_{t-1} 之间没有互动，这时，无条件变化模型比条件变化模型好。不过，艾利森提到，当 Y 的初始值与治疗有任何互动时（例如，一个人因为数学成就测验成绩差而被选去参加数学辅导），条件变化模型可能比无条件变化模型更可取。此外，如果 Y_{t-1} 对 Y 有因果作用，那么条件变化模型可能更合适。这就提出了一个概念性的问题，它有时在讨论条件与无条件变化模型的相对优势时被忽略了，即社会惯性。如戴维斯所言："积极的时间序列依赖或惯性是大多数可预料的社会行为。"（Davis，1994：33）麦金尼斯认为，"保持任何状况的概率增加，当前值在该状况下是一个严格的单调函数"（McGinnis，1968：716）。芬克尔指出，Y 的前值可能会影响 Y 的当前值，这与 Y_{t-1} 对 Y 的影响可能会被无条件变化模型错误地指定一样（Finkel，1995：7）。

滞后内生变量的系数有时称为"稳定系数"。有几种统计上难以区分的解释，最适当的解释必须由概念或理论上的考虑来决定（Davis，1994；Finkel，1995；Kessler & Greenberg，1981；Rogosa，1995）。最常见的解释是控制事先不可测量的成分对 Y 的影响，或是 Y 过去值对当前值的惯性影响。另外，它也可能解释为在同一时间做几件事情。戴维斯指出，稳定的回归系数可能代表以

前的状态或行为对当前状态或行为的效应，以及可测量变量的之前效应和不可测量的变量的效应对因变量 Y 的影响。总之，条件变化模型对惯性效应的估计较为宽松，而且对模型中的其他预测变量影响的估计较为保守。因此，戴维斯(1994：36—37)指出，条件变化模型是很不完美的，它倾向于高估惯性效应并低估其他预测变量对 Y 的影响(Davis, 1994：34—37)。通过观察，干预的影响往往低于统计模型所预测的，这一特性可能是条件变化模型的优点(Davis, 1994：32)。

潜变量增长曲线模型

在线性小组模型中，首要的重点往往是预测或解释个案间一个或多个因变量的差异，然后再描述变化的模式。潜变量增长曲线模型(Bijleveld et al.，1998：第 4 章；Kaplan, 2000：第 8 章；McArdle & Bell, 2000；Stoolmiller, 1995)的重点次序刚好相反，最重要是变化的说明或分析。沿用比莱韦德等人的简化符号(Bijleveld et al.，1998：250)，没有共变项的潜变量增长曲线模型可以写成 $\hat{Y}_t = Z_1 + Z_2 t$，$\hat{Y}_t$ 是 Y 在时间 $t(t=0)$ 时的观测值，$t=0, 1, 2, \cdots, T$，是时间(或年龄)指数，Z_1 是潜变量参数代表 Y 的截距或初始值 $(t=0)$，Z_2 是潜变量参数代表随时间 Y 值增长曲线的增长率或斜率。Z_1 通常代替了 Y 截距——固定参数 α，Z_2 通常代表固定参数 β 作为增长曲线(时间系数)的斜率。共变量可加到模型中，以便解释因变量 Y 或潜变量增长曲线的参数 Z_1 和 Z_2，如果是后者，该模型结构通常与多层增长曲线模型相似，这将稍后讨论。但是，这样的模型可能有点笨重。例如，司徒米勒用了 27 个方程来说明一个 4 期的智力潜变量增长曲线模型的详细规范(Stoolmiller, 1995)。因变量潜变量增长曲线模型的例子还有认知表现、心理健康、酒精和药物滥用，以及其他非法与偏差行为，这些都可在其他著作中找到(Little et al.，2000；Collins & Sayer, 2001)。

在 SEM 潜在增长曲线模型的教科书与实证研究中，时期的数量通常是 2 至 7，4 或 5 波数据是常态(Collins & Sayer, 2001；Gottrnan, 1995；Little et al.，2000)。对于少量的时间序列模型，潜变量增长曲线模型的结果通常与多层增长曲线模型的结果相同或非常相似(Little et al.，2000)。实际上，使用潜变量增长曲线模型会涉及更少的变量，所有没有测量和原则上无法测量的变量

的可靠性都是适当的，每个变量有3至5个指标以及可能很复杂的协方差结构，并且可能包括相关的误差。更多的变量或更多的指标就需要更大的样本量或对模型有更多的限制(这需要理论依据)。比莱韦德等人认为:“一般来说，结构方程模型适用于需要高度识别的理论模型，大量个案只做较少次数的测量。”(Bijleveld, 1998: 267)他们的观察显示，拟合大量时期的结构方程模型，如果个案数量太少，可能变得很复杂，或者得出不适当的答案。比莱韦德等人提议，先拟合模型中的小部分，然后将各小部分结果组合起来，直到得到满意的结果为止。他们的讨论为结构方程模型在追踪数据方面的实际应用提供了合理的意见(Bijleveld, 1998: 265—268)。

质变的描述模型:阶段分析和转换矩阵

质性数据变化的模型通常使用阶段分析，将数据分类到有限的类别，这样，个案可以随时间从一个类别转移到另一个。变化阶段模型关注的是在某一时期(阶段)，一个值(状态)转变到另一个值的概率。对于多分类或多值分类变量的不同转换(在某特定的时期之间，从一个值到另一个值)的概率，每对原始值(区间内的开始状态或数值)和最终值(区间内的最后状态或数值)都要计算，包括原始值与最终值相等的情况。当原始值与最终值相等时，转换概率表示在特定的时间间隔内，状况的稳定性。

社会科学中改变的阶段模型通常是概率性而不是决定性的。阶段转换可用简单转换矩阵来描述，对转换矩阵的性质没有任何假设(例如，Elliott et al., 1989: 179):马尔可夫模型，包括马尔可夫链(Bartholomew, 1973; Bijleveld et al., 1998:第6章);对数线性模型(Bijleveld et al., 1998:第6章;Hout, 1983);潜变量方法包括混合马尔可夫潜在类别模型和潜在转换模式(Collins, 2001; Collins et al., 2000; Langeheine & van de Pol, 1994);生命表模型(Namboodiri & Suchindran, 1987);生存、风险或事件史分析模型(Allison, 1984; Blossfeld et al., 1989; Hosmer & Lemeshow, 1999; Yamaguchi, 1991)。转换矩阵，如马尔可夫模型和对数线性模型，是基于交叉表或列联表值的简单的行百分比，这是将所有个案中的某个变量(列变量)在某一时间的值与该变量(行变

量)后期的值进行比较。有些马尔可夫模型存在吸收状态,即是一旦进入,就不能离开。最常见的吸收状态的例子是死亡。对于至少有一个吸收状态的齐性马尔可夫过程,每个案件最终将进入一个吸收状态,可计算在某特定时期,个案进入吸收状态和其他状态的比例,以及全部或相当比例的个案都进入吸收状态所需要的时间。

埃利奥特等人应用转换矩阵来模拟青春期从不犯罪到犯罪和吸毒的情况(Elliott et al.,1989)。他们用了 5 个阶段(1976—1980 年)和 4 个州(不犯罪、探索犯罪、不严重的固定犯罪模式、严重犯罪;没有使用药物、喝酒、吸食大麻和滥用多种药物)。根据马库斯做的卡氏检定,犯罪的转换矩阵是齐性的,他们从一个时期到另一个时期的变化不超过随机误差。青春期违法矩阵近似一个固定的马尔可夫过程。吸毒的转换矩阵从一个时期到另一个时期会出现显著的差别或非齐性,主要因为更严重的药物滥用的转换模式更有可能出现在青春期后期,而非青春期早期。埃利奥特等人部分地分析了用转换矩阵来描述违法行为的发展模式,他们还利用转换矩阵来检测不同的违法行为的开始和暂停。

多个案长时间序列方法

表 5.2 的右上角是纵贯研究人员的梦想:许多个案和许多时期(实际上,从 ARIMA 模型的角度来看,不必有许多时期,但这更多是为了其他统计方法)。然而,只有两种分析方法列在该表格中:连续时间事件史分析和多层增长曲线模型。两者也列在右下角(很多个案、很少时期)。利用其他方法,如线性追踪样本分析、潜变量增长曲线分析、阶段模型,甚至方差分析、协方差分析、多元方差分析以及共变量(多元协方差分析)模型来分析长时间序列数据会很困难。相反,历史事件分析和多层增长曲线模型更为灵活,可用于长的或短的时间序列数据。

事件史分析

事件史分析(Blossfeld et al.,1989;Hosmer & Lemeshow,1999;Nam-

boodiri & Suchindran, 1987; Yamaguchi, 1991)包括生存和风险分析，都是关联几个时期的数据的回归分析和转换矩阵分析。事件史分析首要关注的是描述、预测并解释质变的时机。事件史分析可用年龄或时间序列作为基本的时间连续，并使用其他时间变量作为一个独立的变量，这样可能检测到历史和发展的趋势。生命表模型包括多状态生命表模型，某种意义上可视为事件史分析的非参数形式。它们分析阶段转换，而且不对转换的时空分布(如事件史分析)作出任何假设，在意义上比事件史分析更灵活，但它们处理大量的自变量却有很大的困难。

对于短时间序列、离散时间(有些学者称其为“分组时间”或“区间检验”，见 Beck et al., 1998; Hosmer & Lemeshow, 1999: 257—269)，事件史分析模型只需要几个时期，可采用 logistic 回归(比例奇数)或互补对数模型(比例危害)。短时序列可能是因为事件发生在离散时间间隔(美国每 4 年一次的总统选举中，选民可选共和党、民主党或独立总统候选人)或粗测时间内(例如，测量相隔时间长，造成每年测量，但这些事件可能在那一年的任何时间发生)。对于较长的系列，通常但不一定涉及更精确的时间测量，无论是半参数考克斯比例风险模型，还是参数事件史分析模型，都可使用(考克斯比例风险模型只是一种情况的比例风险模型；一些参数和离散时间模型也是比例风险模式)。事件史分析被用于描述分析和因果分析，例如，累犯模型(Schmidt & Witte, 1988)、劳动力比率(Blossfeld et al., 1989)、婚姻历史事件(Peters, 1988)以及其他涉及离散状态间的转换事件。

多层增长曲线模型和相关方法

个体或个案内变化的多层增长曲线模型在比莱韦德等人、劳登布什和白克、斯尼德斯和博斯克的专著中都有讨论。劳登布什等人以及斯尼德斯和博斯克用了几章的篇幅介绍了分类因变量(二分、名义、序数、数量)的情况(Raudenbush et al., 2000; Snijders & Bosker, 1999)。基本的多层纵贯数据分析模型涉及两个层面：个人或个案层面(第二层)，这层数据描述个案的特征不随时间变化；观察层面(第一层)，这层重复测量个人特征，包括因变量，这些特征会随

时间而变化。简单的描述性增长曲线模型不包括第二层的预测变量，只有第一层的时间或年龄（或两者）的测量才是预测变量（例如，X_1 =时间）。在这种情况下，时间对因变量的影响可以说是固定的（而不是随机的，即变量）。更复杂的模型可以包括更多的复杂时间函数（例如，二次或三次多项式）和第二层额外的不随时间改变的变量以及第一层的时变变项。总而言之，这两层可以检测固定的个别特点对因变量以及因变量与时间之间的关系，还能检测因变量和第一层其他预测变量随着时间的变化，其中包括检测不同的个人特征是否以及如何产生影响，预测变量和因变量之间的关系是否以及如何随着年龄或时间变化。换句话说，增长曲线的斜率受到个别特点的影响，包括该个案的固定特性和时变的特性。

比莱韦德等人（Bijleveld et al.，1998：第 3 章）也应用重复测量单元方差分析（ANOVA）及协方差（ANCOVA）、多元方差分析（MANOVA）和协方差（MANCOVA）来分析纵贯数据，并提供了很有用的多层模型比较。在实际应用中，重复 ANOVA、ANCOVA、MANOVA 和 MANCOVA 都是多层混合模型（包括多层增长曲线模型）的子集，后者一般更适用于多于 2 个或 3 个时期的纵贯数据分析。重复 ANOVA、ANCOVA、MANOVA 和 MANCOVA 在分析很短系列的量化数据时是最有用的，这种情况在实验与伪实验研究中最常见。多层增长曲线模型更灵活，可以用于量化因变量的定性分析。不像潜变量增长曲线模型，它们可以轻松地处理长时序（大量时期）。事实上，多层增长曲线模型的可靠性随着时期的增加而提高，它们还可以处理通常用潜变量增长曲线模型的短系列数据。当比较这两种方法在较短系列数据中的应用时，其结果往往是相似的（Little et al.，2000）。劳登布什和白克应用了多层增长曲线来模拟认知发展和词汇学习（Raudenbush & Bryk，2002）。劳登布什使用全国青年调查数据和多层增长曲线模型来检测整个生命过程中，对违法行为、结交违法朋友、年龄和性别之间的关系的态度（Raudenbush，1995），并用同样的方法分析了二分因变量的纵贯数据以及 11 岁至 21 岁期间有严重盗窃行为的概率。

结论：纵贯与截面数据和分析

本书比较了纯截面研究与纵贯研究，并界定了纵贯研究的目的。现在总结一下两者的不同。

第一，纵贯研究通常成本较高。如果研究问题或假设可以用截面数据，就不必用纵贯研究去回答研究问题或检验假设。

第二，纵贯研究与截面研究都面对同样的问题：数据质量和充足样本。除此之外，还有其他。虽然有办法解决这些问题，但如果截面研究适合，同样不必用纵贯研究。

第三，截面研究无法解决发展（岁）趋势、历史（期）趋势及世代效应。当这3个类型的效果有可能出现时，研究这种变化就需要用纵贯数据。

第四，历史变迁的描述和分析绝对需要纵贯数据。同时，相对于一般常见的纵贯和截面分析而言，纵贯分析方法，如差分方程模型、ARIMA时序模型、事件史分析，可提供更强大且更详细的历史性变化分析。

第五，虽然可以利用截面特定年龄（或指定阶段）的数据来描述和分析发展趋势，但是这些结果未必能反映出纵贯数据的结果。至于将发展变化视为反映随着年龄或阶段的个人经验的纵贯数据，是因为它们反映了个体内的变化，而不是个体间的差异。

第六，除非有充分的理由，否则（除了众所周知的，如动态过程是非遍历的）应当假定，如果要更有效且更正确地估计社会科学任何动态过程中的参数，就必须使用纵贯数据。

第七，除非回忆期很短或应答条件习惯效应很严重，或者能够证明长期回忆的问题是轻微或不存在的，前瞻追踪样本设计或总人口设计通常比其他纵贯设计好。

第八，检测时间或因果次序应该是检测因果关系中不可分割的一部分。与共变关系（显示在关系强度中）和非虚假关系（显示在关系的持续意义中）一样，时间或因果次序会显示在阶段状态时间序列分析中。格兰杰因果关系或线性

追踪样本的分析就是任何因果关系的重要元素。

鉴于这些结论，截面研究还有什么作用呢？最明显的答案是，截面研究仍然可用于描述特定时间内的变量和关系模式。另外，截面研究的耗费比纵贯研究少，截面设计能够了解动态模型中的假设或对研究问题进行探索或初步的调查。如果关注同一时间、不同年龄的个体之间的差异，而不是推断个体在整个生命过程中随着年龄而发生变化，截面研究是可取的。但是，如果是描述和分析动态变化的过程，那么纵贯研究最终是必不可少的。原则上，纵贯研究能做很多截面研究所不能做的事，相反，没有截面研究能做而纵贯研究不能做的事。

纵贯研究不能解决截面研究的所有问题。它不能解决差劣研究设计、样本量不足或忽略分析方法的假设和限制所带来的问题（相反，很可能会加重问题的严重性）。并非所有的研究问题都要纵贯研究，截面研究也可以做很多。最好把纵贯研究视为社会科学家可用的一个强大的工具。如果研究的问题并不需要纵贯设计，那么使用它将浪费时间、金钱和精力。如果研究问题或假设确实需要纵贯数据和分析，而且能很好地利用，那么结果的质量才能充分地补偿纵贯研究的成本。

注释

[1] 有关使用社会指标的讨论和社会变化指标的介绍，见鲍尔以及谢尔登和摩尔(Bauer, 1966; Sheldon & Moore, 1968)的著作。美国人口普查局列出了美国从殖民时代到1970年的统计数字，卡罗等人列出了20世纪美国众多社会指标的趋势(Caplow et al., 2001)。

[2] 因果模型的基本讨论，见阿舍(Asher, 1983)、布莱洛克(Blalock, 1964)、戴维斯(Davis, 1985)、海斯(Heise, 1975)和皮尔(Pearl, 2000)等人的著作。从哲学角度的讨论，见纳格尔的著作(Nagel, 1961)。对非实验研究的因果推论的争论可在马斯登的著作中(Marsden, 1991)找到，麦克吉姆和特纳(McKim & Turner, 1997)以及谢弗(Shaffer, 1992)的讨论多限于模拟个体内的变化(而不是个体间差异的因果分析)，而且数据涉及少量的时期或截面设计。在"硬"科学领域中(如天文学)，非实验研究的因果推论通常被忽略或轻轻带过，社会科学需要论据和"近实验化"的研究标准(暗示治疗的随机分配和操纵推定的因果变量)。简言之，在我看来，实验和非实验研究提供的是关于因果关系的证据而不是证明，两者都有内在的限制(Babbie, 2001: 226—235; Campbell & Stanley, 1963; Cook & Campbell, 1979)和执行不力的潜在局限性。虽然实验研究较之非实验研究有明确的优势，但是这种优势是在程度上，而不是不同类型。

[3] 有人提出附加标准。例如，坚持一些因果的机制或连接。这些都是相当模糊的标准。在物理科学中，它可能包括"某距离的行动"的拒绝原则，这是一个大量/能量集群影响另一个大量/能量集群，必须有一些接触、粒子或波交换，但量子理论显然暗示某距离的行动。汤姆森写道："量子力学的因果关系是统计，它适用于大多数个体。它的概率通常介于0和1，惯用的解释是，相当一部分人做一件事而另一部分人做别的事。"(Thomsen, 1987: 346)该因果关系的解释和"量子力学不能作出预测个别对象"与社会科学的因果想法一致，某距离的行动似乎排除了因果关系第四个标准的需要(机制或联系)。更深入的讨论请参阅巴比(Babbie, 2001:第3章)、布莱洛克(Blalock 1964、1971)、库克和坎贝尔(Cook & Campbell, 1979:第1、第4章)以及纳格尔(Nagel, 1961)的著作。不同的观点认为，在社会科学调查中，使用"因果关系"这个名词不妥当且不必要(Kerlinger, 1986: 361)。

[4] 全国犯罪调查后来改名为"全国犯罪受害调查"，基于研究调查方法术语的一致性，这本书会使用较早的用词，除了比较两者之外。

[5] 如果 N 是所有时期的个案数量，T 是个案数据的时期数目，那么 $N(T-1)$ 的个案就足够使用某个特定的方法(例如，平均值差异的显着测验或三四个自变量的多元回归)，也可以用集合截面和单一的二三波的纵贯时间序列数据来分析。

[6] 塔里斯为处理纵贯研究的缺失数据提供了很详尽的介绍(Taris, 2000)。艾利森(Allison, 2002)、罗维和德莱尼(Rovine & Delaney, 1990)提供了类似但更深入的介绍。卡瑟波等人(Kasprzyk et al., 1989)和李涛等人(Little et al., 2000)用了几章的篇幅讲解了纵贯研究的缺失数据的问题。列波斯基(Lepkowski, 1989)比较了加权和插补法来调整回应流失，包括对稍后波、跨波以及同波参加者的流失进行插补，并用结合加权和插补法来处理缺失数据。

参考文献

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Ahluwalia, M. S. (1974). "Income inequality: Some dimensions of the problem." In H. B. Chenery, M. S. Ahluwalia, C. L. G. Bell, J. H. Duloy, & R. Jolly(Eds.), *Redistribution with growth: An approach to policy* (pp. 3—37). Oxford, UK: Oxford University Press.

Ahluwalia, M. S. (1976). "Inequality, poverty, and development." *Journal of Development Economics*, *2*, 307—342.

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: Sage.

Allison, P. D. (1990). "Change scores as dependent variables in regression analysis." In C. C. Clogg (Ed.), *Sociological methodology 1990* (Vol. 20, pp. 93—114). Washington, DC: American Sociological Association.

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Appleman, P. (Ed.). (1976). *Thomas Robert Malthus: An essay on the principle of population*. New York: Norton.

Armas, G. C. (2001, March 2.). "Battle brewing over census count." Associated Press.

Asher, H. B. (1983). *Causal modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Babbie, E. (2001). *The practice of social research* (9th ed.). Belmont, CA: Wadsworth.

Baltes, P. B., Cornelius, S. W., & Nesselroade, J. R. (1979). "Cohort effects in developmental psychology." In J. R. Nesselroade & P. B. Baltes(Eds.), *Longitudinal research in the study of behavior and development* (pp. 61—87). New York: Academic Press.

Baltes, P. B., & Nesselroade, J. R. (1979). "History and rationale of longitudinal research." In J. R. Nesselroade & P. B. Baltes(Eds.), *Longitudinal research in the study of behavior and development* (pp. 1—39). New York: Academic Press.

Barnard, J. R., & Krautmann, A. C. (1988). "Population growth among U. S. regions and metropolitan areas: A test for causality." *Journal of Regional Science*, *22*, 103—118.

Bartholomew, D. J. (1973). *Stochastic models for social processes* (2nd ed.). New York: John Wiley.

Bauer, R. A. (Ed.). (1966). *Social indicators*. Cambridge: MIT Press.

Beck, N., & Katz, J. N. (1995). "What to do(and not to do) with time-series cross-section data." *American Political Science Review*, *82*, 634—647.

Beck, N., Katz, J. N., & Tucker, R. (1998). "Taking time seriously: Time-series-cross section analysis with a binary dependent variable." *American Journal of Political Science*, *42*, 1260—1288.

Becker, G. S., Landes, E. M., & Michael, E T. (1977). "An economic analysis of marital instability." *Journal of Political Economy*, *82*, 1141—1187.

Berrueta-Clement, J., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P.

(1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19*. Ypsilanti, MI: HigWScope Educational Research Foundation.

Berry, W. D. (1984). *Nonrecursive causal models*. Beverly Hills, CA: Sage.

Bickman, L., & Rog, D. J. (Eds.). (1998). *Handbook of applied social research methods*. Thousand Oaks, CA: Sage.

Bljleveld, C. J. H., & van der Kamp, L. J. T., with Mooijaart, A., van der Kloot, W. A., van der Leeden, R., & van der Burg, E. (1998). *Longitudinal data analysis: Designs, models, and methods*. London: Sage.

Black, C. E. (1966). *The dynamics of modernization*. New York: Harper & Row.

Blalock, A. B., & Blalock, H. M., Jr. (1982). *An introduction to social research*. Englewood Cliffs, NJ: Prentice Hall.

Blalock, H. M., Jr. (1962). "Four-variable causal models and partial correlations." *American Journal of Sociology*, *62*, 182—194.

Blalock, H. M., Jr. (1964). *Causal inference in nonexperimental research*. New York: Norton.

Blalock, H. M., Jr. (1969). *Theory construction: From verbal to mathematical formulations*. Englewood Cliffs, NJ: Prentice Hall.

Blalock, H. M., Jr. (Ed.). (1971). *Causal models in the social sciences*. Chicago: Aldine.

Blau, P. M., & Duncan, O. D. (1966). *The American occupational structure*. New York: John Wiley.

Blossfeld, H., Hamerle, A., & Mayer, K. U. (1989). *Event history analysis: Statistical theory and application in the social sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Blumstein, A., Cohen, J., Roth, J. A., & Visher, C. A. (Eds.). (1986). *Criminal careers and "career criminals"* (Vols. 1, 2). Washington, DC: National Academy Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bohrnstedt, G. W. (1983). "Measurement." In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 69—121). Orlando, FL: Academic Press.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). San Francisco: Holden-Day.

Brehm, J. (1993). *The phantom respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.

Brown, C., Duncan, G. J., & Stafford, F. P. (1996). "Data watch: The Panel Study of Income Dynamics." *Journal of Economic Perspectives*, *12*, 155—168.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Bulmer, M. G. (1979). *Principles of statistics*. New York: Dover.

Bureau of Justice Statistics. (1992). *Criminal victimization in the United States: 1973—1990 trends*. Washington, DC: U. S. Department of Justice.

Burgess, R. D. (1989). "Major issues and implications of tracing survey respondents." In D.

Kasprzyk, G. Duncan, G. Kalton, & M. P. Singh(Eds.), *Panel surveys* (pp. 52—74). New York: John Wiley.

Caldwell, J. C. (1976). "Toward a restatement of demographic transition theory." *Population and Development Review*, *2*, 321—366.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Cantor, D. (1989). "Substantive implications of longitudinal design features: The National Crime Survey as a case study." In D. Kasprzyk, G. Duncan, G. Kalton, & M. P. Singh(Eds.), *Panel surveys*(pp. 25—51). New York: John Wiley.

Caplow, T., Hicks, L., & Wattenberg, B. J. (2001). *The first measured century: An illustrated guide to trends in America, 1900—2000*. Washington, DC: AEI Press.

Carlson, E. (1979). "Divorce rate fluctuation as a cohort phenomenon." *Population Studies*, *32*, 523—536.

Chilton, R., & Spielberger, A. (1971). "Is delinquency increasing? Age structure and the crime rate." *Social Forces*, *42*, 487—493.

Clarridge, B. R., Sheehy, L. L., & Hauser, T. (1977). "Tracing members of a panel: A 17-year follow-up." In K. F. Schuessler(Ed.), *Sociological methodology 1978* (pp. 185—203). San Francisco: Jossey-Bass.

Collins, C., Given, B., & Berry, D. (1989). "Longitudinal studies as intervention." *Nursing Research*, *32*, 251—253.

Collins, L. (2001). "Reliability for static and dynamic categorical latent variables: developing measurement instruments based on a model of the growth process." In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change*(pp. 273—288). Washington, DC: American Psychological Association.

Collins, L. M., Hyatt, S. S., & Graham, J. W. (2000). "Latent transition analysis as a way of testing models of stage-sequential change in longitudinal data." In, T. D. Little, K. U. Schnabel, & J. Baumert(Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specijic examples*(pp. 147—161). Mahwah, NJ: Lawrence Erlbaum.

Collins, L. M., & Sayer, A. G. (Eds.). (2001). *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafing the standardized questionnaire*. Beverly Hills, CA: Sage.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cordray, S., & Polk, K. (1983). "The implications of respondent loss in panel studies of deviant behavior." *Journal of Research in Crime and Delinquency*, *22*, 214—242.

Covey, H. C., & Menard, S. (1987). "Trends in arrests among the elderly." *The Gerontologist*, *22*, 666—672.

Covey, H. C., & Menard, S. (1988). "Trends in elderly criminal victimization from 1973— 1984." *Research on Aging*, *12*, 329—341.

Cromwell, J. B., Hannan, M. J., Labys, W. C., & Terraza, M. (1994). *Multivariate tests for time series models*. Thousand Oaks, CA: Sage.

Cronbach, L. J., & Furby, L. (1970). "How should we measure change-Or should we?" *Psychological Bulletin*, *72*, 68—80.

Davies, R. B. (1994). "From cross-sectional to longitudinal analysis." In A. Dale & R. B. Davies (Eds.), *Analyzing social and political change: A casebook of methods* (pp. 20—40). London: Sage.

Davies, R. B., & Pickles, A. R. (1985). "Longitudinal versus cross-sectional methods for behavioral research: A first-round knockout." *Environment and Planning A*, *12*, 1315—1329.

Davis, J. A. (1985). *The logic of causal order*. Beverly Hills, CA: Sage.

Davis, J. A., & Smith, T. W. (1992). *The NORC General Social Survey: A user's guide*. Newbury Park, CA: Sage.

Davis, K. (1963). "The theory of change and response in modem demographic history." *Population Index*, *22*, 345—366.

Dempster-McClain, D., & Moen, P. (1998). "Finding respondents in a follow-up study." In J. Z. Giele & G. H. Elder, Jr. (Eds.), *Methods of life course research* (pp. 128—151). Thousand Oaks, CA: Sage.

Easterlin, R. A. (1987). *Birth andfortune* (2nd ed.). Chicago: University of Chicago Press. Elliott, D. S., Huizinga, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use*. Beverly Hills, CA: Sage.

Elliott, D. S., Huizinga, D., & Menard, S. (1989). *Multiple problem youth: Delinquency, substance use, and mental health problems*. New York: Springer-Verlag.

Federal Bureau of Investigation. (annual). *Uniform crime reports*. Washington, DC: Government hinting Office.

Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage.

Firebaugh, G. (1980). "Cross national versus historical regression models: Conditions of equivalence in comparative analysis." *Comparative Social Research*, *2*, 333—344.

Fowler, E(1998). "Design and evaluation of survey questions." In L. Bickman & D. J. Rog(Eds.), *Handbook of applied social research methods* (pp. 343—374). Thousand Oaks, CA: Sage.

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-Demarco, L. (1988). "The life history calendar: A technique for collecting retrospective data." In C. Clogg(Ed.), *Sociological methodology 1988* (pp. 37—68). Washington, DC: American Sociological Association.

Freeman, D. (1983). *Margaret Mead and Samoa: The making and unmaking of an anthropological myth*. Cambridge, MA: Harvard University Press.

Giele, J. Z., & Elder, G. H., Jr. (Eds.). (1998). *Methods of life course research: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Gist, N. P., & Fava, S. E(1974). *Urban society* (6th ed.). New York: Crowell/Harper & Row.

Glenn, N. (1976). "Cohort analysts' futile quest: Statistical attempts to separate age, period, and cohort effects." *American Sociological Review*, *41*, 900—904.

Glenn, N. (1977). *Cohort analysis*. Beverly Hills, CA: Sage.

Gold, M. , & Reimer, D. J. (1975). "Changing patterns of delinquent behavior among Americans 13 through 16 years old: 1967—1972." *Crime and Delinquency Literature*, *2*, 483—577.

Gottman, J. M. (Ed.). (1995). *The analysis of change*. Mahwah, NJ: Lawrence Erlbaum.

Graetz, B. (1987). "Cohort changes in educational inequality." *Social Science Research*, *12*, 329—344.

Graham, J. W. , & Hofer, S. M. (2000). "Multiple imputation in multivariate research." In T. D. Little, K. U. Schnabel, & J. Baumert(Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201—218). Mahwah, NJ: Lawrence Erlbaum.

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and crossspectral methods." *Econometrica*, *32*, 424—438.

Greenberg, D. F. (1985). "Age, Crime, and social explanation." *American Journal of Sociology*, *91*, 1—21.

Grusky, D. B. (2001). *Stratification: class, race, and gender in sociological perspective* (2nd ed.). Boulder, CO: Westview.

Hamblin, R. L. , Jacobsen, R. B. , & Miller, J. L. L. (1973). *A mathematical theory of social change*. New York: John Wiley.

Hartford, R. B. (1984). "The case of the elusive infant mortality rate." *Population Today*, *2*, 6—7.

Hayduk, L. A. (1987). *Structural modeling with WSREL: Essentials and advances*. Baltimore: Johns Hopkins University Press.

Heise, D. R. (1975). *Causal analysis*. New York: John Wiley.

Henry, B. , Moffitt, T. E. , Caspi, A. , Langley, J. , & Silva, P. A. (1994). "On the 'remembrance of things past': A longitudinal evaluation of the retrospective method." *Psychological Assessment*, *2*, 92—101.

Heyns, B. (1978). *Summer learning and the effects of schooling*. New York: Academic Press.

Hill, M. S. (1992). *The Panel Study of Income Dynamics: A user's guide*. Newbury Park, CA: Sage.

Hobcraft, J. , Menken, J. , & Preston, S. (1982). "Age, period, and cohort effects in demography: A review." *Population Index*, *42*, 4—43.

Hogan, H. , & Robinson, G. (2000). *What the Census Bureau's coverage evaluation programs tell us about the differential undercount*. Washington, DC: U. S. Department of Commerce, Bureau of the Census.

Hosmer, D. W. , Jr. , & Lemeshow, S. (1999). *Applied survival analysis: Regression modeling of time to event data*. New York: John Wiley.

Hout, M. (1983). *Mobility tables*. Beverly Hills, CA: Sage.

Hout, M. , Manza, J. , & Brooks, C. (1999). "Classes, unions, and the realignment of U. S. presidential voting, 1952—1992." In G. Evans(Ed.), *The end of classpolitics: Class voting in comparative perspective* (pp. 83—95). Oxford, UK: Oxford University Press.

Huckfeldt, R. R. , Kohfeld, C. W. , & Likens, T. W. (1982). *Dynamic modeling: An introduction*.

Beverly Hills, CA: Sage.

Huizinga, D. H., Menard, S., & Elliott, D. S. (1989). "Delinquency and drug use: Temporal and developmental patterns." *Justice Quarterly*, *2*, 419—455.

Inglehart, R. (1997). *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*. Princeton, NJ: Princeton University Press.

Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications*. San Francisco: Holden-Day.

Johnston, L. D., Bachman, J. G., & O'Malley, P. M. (annual). *Monitoring the future: Questionnaire responses from the nation's high school seniors*. Ann Arbor, MI: Institute for Social Research.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 2: A guide to the program and applications* (2nd ed.). Chicago: SPSS.

Kalton, G., Kasprzyk, D., & McMillen, D. B. (1989). "Nonsampling errors in panel surveys." In D. G. Kasprzyk, G. Duncan, G. Kalton, & M. P. Singh (Eds.), *Panel surveys* (pp. 249—270). New York: John Wiley.

Kandel, D. B. (1975). "Stages of adolescent involvement in drug use." *Science*, *192*, 912—914.

Kandel, D. B., & Faust, R. (1975). "Sequence and states in patterns of adolescent drug use." *Archives of General Psychiatry*, *32*, 923—932.

Kandel, D. B., & Logan, J. A. (1984). "Patterns of drug use from adolescence to young adulthood: I. Periods of risk for initiation, continued use, and discontinuation." *American Jounal of Public Health*, *72*, 660—666.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Kelling, G. L., Pate, T., Dieckman, D. E., & Brown, C. E. (1974). *The Kansas City Preventative Patrol Experiment: A summary report*. Washington, DC: Police Foundation.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart & Winston.

Kessler, R. C., & Greenberg, D. E. (1981). *Linear panel analysis: Models of quantitative change*. New York: John Wiley.

Kim, K. H., & Roush, F. W. (1980). *Mathematics for social scientists*. New York: Elsevier.

Knoke, D., & Hout, M. (1976). "Reply to Glenn." *American Sociological Review*, *41*, 905—908.

Kraemer, H. C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.

Lagrange, R. L., & White, H. R. (1985). "Age differences in delinquency: A test of a theory." *Criminology*, *22*, 19—45.

Langeheine, R., & van de Pol, E (1994). "Discrete-time mixed Markov latent class models." In A. Dale & R. B. Davies (Eds.), *Analyzing social and political change: A casebook of methods* (pp. 170—197). London: Sage.

Lehnen, R. G., & Skogan, W. G. (Eds.). (1981). *The National Crime Survey: Working papers: Vol. I. Current and historical perspectives*. Washington, DC: U. S. Department of Justice.

Lepkowski, J. M. (1989). "Treatment of wave nonresponse in panel surveys." In D. G. Kasprzyk, G. Duncan, G. Kalton, & M. P. Singh(Eds.), *Panel surveys*(pp. 348—374). New York: John Wiley.

Lewis, O. (1951). *Life in a Mexican village: Tepoztlan restudied.* Urbana: University of Illinois Press.

Liker, J. K., Augustyniak, S., & Duncan, G. J. (1985). "Panel data and models of change: A comparison of first difference and conventional two-wave models." *Social Science Research*, *12*, 80—101.

Little, R. J. A., & Su, H. (1989). "Item nonresponse in panel surveys." In D. Kasprzyk, G. J. Duncan, G. Kalton, & M. P. Singh(Eds.), *Panel surveys*(pp. 400—425). New York: John Wiley.

Little, T. D., Schnabel, K. U., & Baumert, J. (Eds.). (2000). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples.* Mahwah, NJ: Lawrence Erlbaum.

Lloyd, L., Armour, P. K., & Smith, R. J. (1987). "Suicide in Texas: A cohort analysis of trends in suicide rates, 1945—1980." *Suicide and Life-Threatening Behavios 12*, 205—217.

Mahajan, V., & Peterson, R. A. (1985). *Models for innovation diffusion.* Beverly Hills, CA: Sage.

Markus, G. B. (1979). *Analyzing panel data.* Beverly Hills, CA: Sage.

Marsden, P. V. (Ed.). (1991). *Sociological methodology 1991.* Washington, DC: American Sociological Association.

Martin, E. (1983). "Surveys as social indicators: Problems in monitoring trends." In P. H. Rossi, J. D. Wright, & A. B. Anderson(Eds.), *Handbook of survey research*(pp. 677—743). Orlando, FL: Academic Press.

Mason, K. O., Mason, W. M., Winsborough, H. H., & POOLE, W. K. (1973). "Some methodological issues in cohort analysis of archival data." *American Sociological Review*, *32*, 242—258.

Mason, W. M., Mason, K. O., & Winsborough, H. H. (1976). "Reply to Glenn." *American Sociological Review*, *41*, 904—905.

Mauldin, W. P., & Berelson, B. (1978). "Conditions of fertility decline in developing countries: 1965—1975." *Studies in Family Planning*, *2*, 89—145.

McArdle, J. J., & Bell, R. Q. (2000). "An introduction to latent growth curve models for developmental data analysis." In T. D. Little, K. U. Schnabel, & J. Baumert(Eds.), *Modeling longitudinal and multilevel data*(pp. 69—107). Mahwah, NJ: Lawrence Erlbaum.

McCord, J. (1983). "A longitudinal study of aggression and antisocial behavior." In K. T. Van Dusen & S. A. Mednick(Eds.), *Prospective studies of crime and delinquency*(pp. 269—275). Boston: Kluwer-Nijhoff.

McGinnis, R. (1968). "A stochastic theory of social mobility." *American Sociological Review*, *23*, 712—722.

McKeown, T. (1976). *The modem rise of population.* London: Edward Arnold.

McKeown, T., & Record, R. (1962). "Reasons for the decline of mortality in England and Wales during the 19th century." *Population Studies*, *12*, 94—122.

McKim, V. R., & Turner, S. P. (Eds.). (1997). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences.* Notre Dame, IN: Notre Dame University Press.

McNeill, W. H. (1976). *Plagues and peoples.* New York: Anchor/Doubleday.

Mead, M. (1928). *Coming of age in Western Samoa: A psychological study of primitive youth for western civilization.* New York: William Morrow.

Menard, S. (1983, April). *Reliability issues in international comparisons of inequality of income.* Paper presented at the annual meeting of the Western Social Science Association, San Diego, CA.

Menard, S. (1986). "A research note on international comparisons of inequality of income." *Social Forces*, *2*, 778—793.

Menard, S. (1987a). "Fertility, development, and family planning, 1970—1982: An analysis of cases weighted by population." *Studies in Comparative International Development*, *22*, 103—127.

Menard, S. (1987b). "Short term trends in crime and delinquency: A comparison of UCR, NCS, and self-report data." *Justice Quarterly*, *2*, 455—474.

Menard, S., & Elliott, D. S. (1990a). "Longitudinal and cross-sectional data collection and analysis in the study of crime and delinquency." *Justice Quarterly*, *2*, 11—55.

Menard, S., & Elliott, D. S. (1990b). "Self-reported offending, maturational reform, and the Easterlin hypothesis." *Journal of Quantitative Criminology*, *2*, 237—267.

Menard, S., & Elliott, D. S. (1994). "Delinquent bonding, moral beliefs, and illegal behavior: A three-wave panel model." *Justice Quarterly*, *11*, 173—188.

Menard, S., Elliott, D. S., & Huizinga, D. (1989). *The dynamics of deviant behavior: A national survey progress report* (National Youth Survey Report No. 49). Boulder, CO: Institute of Behavioral Science.

Menard, S., & Huizinga, D. (1989). "Age, period, and cohort size effects on self-reported alcohol, marijuana, and polydrug use: Results from the National Youth Survey." *Social Science Research*, *12*, 174—194.

Menard, S., & Mihalic, S. (2001). "The tripartite conceptual framework in adolescence and adulthood: Evidence from a national sample." *Journal of Drug Issues*, *31*, 905—938.

Mensch, B. S., & Kandel, D. B. (1988). "Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects." *Public Opinion Quarterly*, *52*, 100—124.

Murray, G. F., & Erickson, P. G. (1987). "Cross-sectional versus longitudinal research: An empirical comparison of projected and subsequent criminality." *Social Science Research*, *12*, 107—118.

Nagel, E. (1961). *The structure of science.* New York: Harcourt, Brace, & World.

Namboodiri, K., & Suchindran, C. M. (1987). *Life table techniques and their applications.* Or-

lando, FL: Academic Press.

Newcomb, M. D. , & Bentler, P. M. (1988). *Consequences of adolescent drug use: Impact on the lives of young adults*. Newbury Park, CA: Sage.

Notestein, F. W. (1945). "Population: The long view." In T. W. Schultz(Ed.), *Food for the world*(pp. 26—57). Chicago: University of Chicago Press.

Ostrom, C. W. , Jr. (1990). *Time series analysis: Regression techniques*. Newbury Park, CA: Sage.

Palmore, E. (1978). "When can age, period, and cohort be separated?" *Social Forces*, *52*, 282—295.

Patterson, G. R. (1995). "Orderly change in a stable world: The antisocial trait as a chimera." In J. M. Gottman(Ed.), *The analysis of change*(pp. 83—101). Mahwah, NJ: Lawrence Erlbaum.

Payne, C. , Payne, J. , & Heath, A. (1994). "Modelling trends in multi-way tables." In A. Dale & R. B. Davies(Eds.), *Analyzing social and political change: A casebook of methods* (pp. 43—74). London: Sage.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Peters, H. E. (1988). "Retrospective versus panel data in analyzing lifecycle events." *Journal of Human Resources*, *22*, 488—573.

Piaget, J. (1948). *The moral judgment of the child*. New York: Free Press.

Piaget, J. (1951). *The child's conception of the world*. New York: Humanities Press.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International University Press.

Platt, J. R. (1964). "Strong inference." *Science*, *142*, 347—353.

Plewis, I. (1985). *Analyzing change: Measurement and explanation using longitudinal data*. Chichester, UK: Wiley.

Population Reference Bureau. (1989). "Speaking graphically: When did you say you were born, Miss?" *Population Today*, *2*, 2.

President's Commission on Law Enforcement and the Administration of Justice. (1967). *The challenge of crime in a free society*. Washington, DC: Government Printing Office.

Raudenbush, S. W. (1995). "Hierarchical linear models to study the effects of social context on development." In J. M. Gottman(Ed.), *The analysis of change* (pp. 165—201). Mahwah, NJ: Lawrence Erlbaum.

Raudenbush, S. W. (2001). "Toward a coherent framework for comparing trajectories of individual change." In L. M. Collins & A. G. Sayer(Eds.), *New methods for the analysis of change* (pp. 35—64). Washington, DC: American Psychological Association.

Raudenbush, S. W. , & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. , Bryk, A. , Cheong, Y. F. , & Congdon, R. (2000). *HLM 2: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.

Redfield, R. (1930). *Tepoztlan: A Mexican village*. Chicago: University of Chicago Press.

Rennison, C. M. (2000). *Criminal victimization 1992: Changes 1998—1999 with trends* 1993—1999. Washington, DC: U. S. Department of Justice.

Richardson, L. F. (1960). *Arms and insecurity*. Pittsburgh: Boxwood Press.

Robey, B. (1989). "Two hundred years and counting: The 1990 census." *Population Bulletin*, *4* (*1*), 1—43.

Rodgers, W. L. (1982a). "Estimable functions of age, period, and cohort effects." *American Sociological Review*, *42*, 774—787.

Rodgers, W. L. (1982b). "Reply to comment by Smith, Mason, and Fienberg." *American Sociological Review*, *42*, 793—796.

Rogosa, D. (1995). "Myths and methods: 'Myths about longitudinal research' plus supplemental questions." In J. M. Gottman(Ed.), *The analysis of change* (pp. 3—66). Mahwah, NJ: Lawrence Erlbaum.

Roitberg, T., & Menard, S. (1995). "Adolescent violence: A test of integrated theory." *Studies on Crime and Crime Prevention*, *2*, 177—196.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Newbury Park, CA: Sage.

Rostow, W. W. (1960). *The stages of economic growth: A non-communist manifesto*. Cambridge, UK: Cambridge University Press.

Rovine, M. J., & Delaney, M. (1990). "Missing data estimation in developmental research." In A. von Eye(Ed.), *Statistical methods in longitudinal research: Vol. I. Principles and structuring change* (pp. 35—79). Boston: Academic Press.

Rubin, Z., & Mitchell, C. (1978). "Couples research as couples counseling: Some unintended effects of studying close relationships." *American Psychologist*, *31*, 17—25.

Rutter, M., Maughan, B., Pickles, A., & Simonoff, E. (1998). "Retrospective recall recalled." In R. B. Cairns, L. R. Bergman, & J. Kagan(Eds.), *Methods and models for studying the individual* (pp. 219—242). Thousand Oaks, CA: Sage.

Ryoder, N. B. (1965). "The cohort as a concept in the study of social change." *American Sociological Review*, *32*, 843—861.

Sanders, D., & Ward, H. (1994). "Time-series techniques for repeated cross-section data." In A. Dale & R. B. Davies(Eds.), *Analyzing social and political change* (pp. 201—223). London: Sage.

Sayrs, L. W. (1989). *Pooled time series analysis*. Newbury Park, CA: Sage.

Schmidt, P., & Witte, A. D. (1988). *Predicting recidivism using survival models*. New York: Springer-Verlag.

Schoenberg, R. (1977). "Dynamic models and cross-sectional data: The consequences of dynamic misspecification." *Social Science Research*, *2*, 133—144.

Schweinhart, L. J., & Weikart, D. P. (1980). *Young children grow up: The effects of the Perry Preschool Program on youths through age 15*. Ypsilanti, MI: High/Scope Educational Research Foundation.

Shaffer, J. P. (Ed.). (1992). *The role of models in nonexperimental social science: Two debates.* Washington, DC: American Educational Research Association and American Statistical Association.

Sheldon, E. B. , & Moore, W. E. (Eds.). (1968). *Indicators of social change: Concepts* and *measurements.* New York: Russell Sage.

Shryock, H. S. , Siegel, J. , & Associates. (1976). *The methods and materials of demography* (Condensed ed. by E. G. Stockwell). New York: Academic Press.

Simon, H. J. (1954). "Spurious correlation: A causal interpretation. " *Journal of the American Statistical Association*, *42*, 467—479.

Sims, C. A. (1972). "Money, income, and causality. " *American Economic Review*, *62*, 540—552.

Skogan, W. G. (1976). "The victims of crime: Some national survey findings. " In A. L. Guenther(Ed.), *Criminal behavior and social systems* (2nd ed. , pp. 131—148). Chicago: Rand McNally.

Smith, H. L. , Mason, W. M. , & Fienberg, S. E. (1982). "More chimeras of the ageperiod-cohort accounting framework: Comment on Rodgers. " *American Sociological Review*, *42*, 787—793.

Sijders, T. , & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Sorenson, A. M. , Brownfield, D. , & Carlson, V. (1989). "Adult reports of juvenile delinquency: A research note on the reliability of a retrospective design. " *Sociological Spectrum*, *2*, 227—237.

Stoolmiller, M. (1995). "Using latent growth curve models to study developmental processes. " In J. M. Gomnan(Ed.), *The analysis of change* (pp. 103—138). Mahwah, NJ: Lawrence Erlbaum.

Stoolmilier, M. , & Bank, L. (1995). "Autoregressive effects in structural equation models: We see some problems. " In J. M. Gottman(Ed.), *The analysis of change* (pp. 261—276). Mahwah, NJ: Lawrence Erlbaurn.

Taris, T. W. (2000). *A primer in longitudinal data analysis.* London: Sage.

Thomlinson, R. (1976). *Population dynamics: Causes and consequences of world population change* (2nd ed.). New York: Random House.

Thompson, W. S. (1929). "Population. " *American Journal of Sociology*, *32*, 959—975.

Thomsen, D. E. (1987). "In the beginning was quantum mechanics: Cosmologists take a chance on a quantum universe. " *Science News*, *131*, 346—347.

Thornberry, T. P. , Bjerregaard, B. , & Miles, W. (1993). "The consequences of respondent attrition in panel studies: A simulation based on the Rochester Youth Development Study. " *Journal of Quantitative Criminology*, *2*, 127—158.

Tolnay, S. E. , & Christenson, R. L. (1984). "The effects of social setting and family planning programs on recent fertility declines in developing countries: A reassessment. " *Sociology and Social Research*, *62*, 72—89.

Tsui, A. O. , & Bogue, D. J. (1978). "Declining world fertility: Trends, causes, implications. " *Population Bulletin*, *33*(*4*), 2—55.

U. S. Bureau of the Census. (1975). *Historical statistics of the United States: Colonial times to 1970* (Bicentennial ed.). Washington, DC: Government Printing Office.

van de Walle, E. , & Knodel, J. (1980). "Europe's fertility transition: New evidence and lessons for today's developing world." *Population Bulletin*, *34*(*6*), 2—40.

Vigderhous, G. (1977). "Forecasting sociological phenomena: Application of Box-Jenkins methodology to suicide rates." In K. F. Schuessler(Ed.), *Sociological methodology 1978* (pp. 20—51). San Francisco: Jossey-Bass.

Wall, W. D. , & Williams, H. L. (1970). *Longitudinal studies and the social sciences*. London: Heinemann.

Walton, M. A. , Ramanathan, C. S. , & Reischl, T. M. (1998). "Tracking substance abusers in longitudinal research: Understanding follow-up contact difficulty." *American Journal of Community Psychology*, *22*, 233—253.

Wei, W. W. S. (1990). *Time series analysis: Univariate and multivariate methods*. Redwood City, CA: Addison-Wesley.

Weikart, D. P. , Bond, J. T. , & McNeil, J. T. (1978). *The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade*. Ypsilanti, MI: HighIScope Educational Research Foundation.

Weis, J. G. (1986). "Issues in the measurement of criminal careers." In A. Blumstein, J. Cohen, J. A. Roth, & C. A. Visher(Eds.), *Criminal careers and "career criminals"* (Vol. 11, pp. 1—51). Washington, DC: National Academy Press.

Wetzel, R. D. , Reich, T. , Murphy, G. E. , Province, M. , & Miller, J. P. (1987). "The changing relationship between age and suicide rates: Cohort effect, period effect, or both?" *Psychiatric Developments*, *2*, 174—218.

Williams, J. , & Gold, M. (1972). "From delinquent behavior to official delinquency." *Social Problems*, *22*, 209—229.

Wofford, S. (1989). *A preliminary analysis of the relationship between employment and delinquency/crime for adolescents and young adults* (National Youth Survey Report No. 50). Boulder, CO: Institute of Behavioral Science.

Wright, R. E. (1989). "The Easterlin hypothesis and European fertility rates." *Population and Development Review*, *12*, 107—122.

Wright, R. E. , & Maxim, P. S. (1987). "Canadian fertility trends: A further test of the Easterlin hypothesis." *Canadian Review of Sociology and Anthmpology*, *22*, 339—357.

Yaffee, R. (with McGee, M.). (2000). *An introduction to time series analysis and forecasting with applications of SAS and SPSS*. San Diego: Academic Press.

Yamaguchi, K. (1991). *Event history analysis*. Newbury Park, CA: Sage.

Yamaguchi, K. , & Kandel, D. B. (1984a). "Patterns of drug use from adolescence to young adulthood: Ⅱ. Sequences of progression." *American Journal of Public Health*, *72*, 668—672.

Yamaguchi, K. , & Kandel, D. B. (1984b). "Patterns of drug use from adolescence to young adulthood: Ⅲ. Predictors of progression." *American Journal of Public Health*, *72*, 673—681.

Zeller, R. A. , & Carmines, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge, UK: Cambridge University Press.

译名对照表

age-specification	特定年龄
ANCOVA	协方差分析
ANOVA	方差分析
autoregression time series	自回归时间序列
causal analysis of differences	差异因果分析
change scores	改变值
cohort membership	世代成员
collinearity	共线性
concurrent validity	同时效度
Conditional Change Model	条件变化模型
Constant Coefficients Model	常系数模型
contemporaneous	同时发生
Cox Proportional Hazards Model	考克斯比例风险模型
deterministic model	决定模型
difference equation	差分方程
Differential Equation Models	微分方程模型
discriminant analysis	判别分析
dummy-variable regression analysis	虚拟变量回归分析
econometric time series	计量经济时间序列
endogenous variable	内生变量
ergodic	遍历
evaluation research	评估研究
event cohort	事件世代
event history analysis	事件史分析
exogenous variable	外生变量
extrapolation	外推法
family of orientation	原生家庭
family of procreation	再生家庭
Granger Causality	格兰杰因果关系
imputation	插补法
interpolation	内推法
interval measure	定距量度
interval scale	定距尺度
lagged Endogenous Variable	滞后内生变量
Latent Growth Curve Model	潜在增长曲线模型

left-hand censoring	左侧删失
life history calendars	生活史日历
linear dependence	线性依赖
linear panel analysis	线性追踪分析
listwise deletion	成列删除
loglinear model	对数线性模型
longitudinal research	纵贯研究
Malthusian Theory	马尔萨斯理论
MANOVA	多元方差分析
Markov Chains	马尔可夫链
measurement interval	测量区间
multiple imputation	多元插补法
multi-year multi-cohort design	多年多世代设计
nonergodic	非遍历
panel conditioning	追踪样本的条件习惯效应
Parametric Event History Analysis Model	参数事件史分析模型
prewhitening	预白噪音化
probabilistic	概率
prospective panel study	前瞻式追踪样本设计
random coefficients	随机系数
raw gain	原增益
regression imputation	回归插补
repeated cross-sectional design	重复截面设计
residual gain	残差增益
retrospective panel design	回顾追踪样本设计
reverse telescoping	逆伸缩
revolving panel design	旋转追踪样本设计
secondary analysis	二次分析
simple autoregressive	简单自回归
social inertia	社会惯性
spectral analysis	光谱分析
Stacked Model	堆栈模式
stage-state analysis of temporal	阶段性时间分析
Stochastic Time Series Model	随机时间序列模型
Structural Equation Modelling	结构方程模型
survival and hazard analysis	生存和风险分析
synthetic cohort	综合世代

telescoping	伸缩
test-retest reliability	再测信度
time series analysis	时间序列分析
Time Series Regression (TSR)	时间序列回归
time-ordered cross-sectional design	时间序列截面设计
time-varying covariates	时变变量
total population design	总人口设计
Unconditional Change Model	无条件变化模型
weighting	加权
white noise process	白噪音过程

分析重复调查数据

作 者 简 介

格伦·菲尔鲍(Glenn Firebaugh)

毕业于美国印第安纳大学布卢明顿分校,现在是美国宾夕法尼亚州立大学社会学系教授和该大学人口研究所的资深科学家。他利用重复调查数据来研究美国社会变迁的人口学根源,并积累跨国数据以研究决定第三世界国家大众福利的因素。目前,他正在研究用于分解不平等变化的方法,并试图运用这些方法来分解不同国家收入不平等的变化。

译 者 简 介

叶 华

现为香港科技大学社会科学部博士候选人,研究方向为社会分层与流动、经济社会学、教育社会学。

序

调查研究是一项正在迅速兴起的课题。一些有价值的介绍实际上假定研究者愿意进行单次调查，成千上万的调查都是这么进行的。在20世纪末，我们发现很多"单次"调查实际上被重复了许多次。重复调查，用菲尔鲍的话来说，就是"向不同的样本询问同样的问题"。众所周知的例子有美国的综合社会调查、全国选举研究和全国健康访问调查。在相当长的一段时间里，这些调查定期对新抽取的全国样本询问相同的问题。举例来说，从1952年起，美国全国选举研究在每次议会和总统选举时都询问大量有关政治态度和行为的问题。

这种重复调查给研究者提供了更多的机会，把社会作为整体来研究它的变迁。它们允许研究者转变关注点，从对个人层次的微观过程的研究，转为对整体层次的宏观过程的研究。菲尔鲍详细解说了用这种方式研究社会变迁的不同方法。为了提供一个一般化的理论取向，首先他尽力解决了怎么分解世代、时期和年龄作用。例如，有人发现，从前越南战争的抗议者对战争的支持度增长了。为什么会这样？这是一种时期作用吗？例如，社会整体上对战争的容忍度提高了？或者这是一种年龄作用吗？年长化会让人更倾向于接受战争？抑或这是一种世代效应，例如，出生在1950年之后的人是否对战争的支持度更高？由于识别上的困难，这些作用很难区分。在注重理论的背景下，菲尔鲍对克服这个困难的方法进行了深入的讨论。

被重复进行过的调查通常允许对总体趋势进行有意义的研究。举例来说，假设在一系列调查中，我们用同样的方法对公众支持战争的态度进行了测量，分析变迁的研究者可以绘图来展现该趋势。此外，也可以对社会中的不同群体进行比较，例如，比较越南战争的抗议者和非抗议者。菲尔鲍用真实数据展示了如何估计群体在趋势上的差异，并令人信服地探讨了年轻的和年长的美国人在社会开支态度上的差别。其总体趋势可以被分解，即把总变迁归因于个人的变化或者人口群体的变化。这被称为"近似分解"，可以通过回归或者代数的方法进行。作为一个例子，菲尔鲍分解了从1972年到1984年反对歧视黑人的趋势的变化，结果表明，相对于个人变化的影响来说，总变化稍微更多地来自世代更替的影响。

对总体趋势的分解可以作为解释性而非描述性的目标。用回归的形式,可以把总体趋势分解为因果性的组成部分,它包括截距的变化、自变量本身的变化和自变量斜率的变化。为了展示更一般的分解方法,菲尔鲍分析了美国选举投票率的变化趋势,并提出了有益的警告且进行了解释,反对机械地运用这个方法。

与研究总体变化相对应,倒数第二章用参数变化模型研究重复调查,此时在个人层次上,自变量 X 对因变量 Y 的作用随时间而变化。重复调查被合并或者说累积成一个数据用于分析。作者细心地选择了美国选举研究数据作为例子,研究影响政党认同的主要因素的变化,例如地区、教育、种族和阶级。

总的来说,作者讨论了重复调查的 4 种基本用处:描述、分解、解释总体趋势,以及对个别参数变化的估计。对重复调查的系统性运用将大大地扩展研究变迁的可能性。菲尔鲍的这篇论述清晰、独一无二的出版物是这个迅速发展的研究领域的一个无价的指引。

迈克尔·S. 刘易斯-贝克

第 1 章 | 导论

重复调查：相同的问题，不同的样本

重复调查即向不同样本的受访者提问相同的问题。因为新的样本在各个不同的测量时期被挑选出来，所以重复调查设计的另一个名称是“重复截面设计”(Menard, 1991)。有一些调查以一定的时间间隔(通常是每月、每季、每年或每两年)重复进行，邓肯和卡顿称之为“定期调查”(Duncan & Kalton, 1987)。其他调查，例如选举前的民意调查，其重复是临时性的。就研究社会变迁而言，亦即本书的重点，定期调查是最容易分析的。

要区分重复调查设计(跨时间的不同样本)和固定样本研究设计(多次调查相同的受访者)，重要的是谨记术语“固定样本”和“重复调查”指的是抽样设计的实质。有些固定样本研究定期加入新样本，有些重复调查包含固定样本，所以术语“重复调查”不一定暗示每次调查都是全新的样本。再访以前的受访者通常会比抽取和访问新的受访者要容易和便宜，尤其是当再访可以通过电话完成的时候。再访使得研究个体随时间的变化成为可能，因为这些个体可以被追踪。

全国性的重复调查已经成为过去 20 多年美国社会科学研究的主要工具。在这些重复调查中，最广为人知的也许就是美国综合社会调查(GSS)，一项每年(1994 年后改为每两年)进行的、涵盖 48 个接壤州①、非制度化的②成年人总体的面访。从 1972 年的首次调查算起，已经有超过 32000 位受访者回答了大约

---

① 即不包括阿拉斯加州和夏威夷。——译者注

② 即不包括军队、学校等制度化组织。——译者注

1500 个涵盖了态度、信仰和行为的众多不同问题。如此搜集而来的数据已经成为社会科学的全国性资源。

在过去的 20 多年中，综合社会调查无疑已经成为社会科学领域除美国人口普查以外被使用得最多的数据。此外，综合社会调查也被广泛用于教学。已有专门的软件被开发出来用于教授社会学专业学生使用综合社会调查，据估计，每年有超过 10 万个社会学专业的学生在课堂中使用综合社会调查。

虽然综合社会调查是“唯一的保证问题会重复的美国调查”(Davis & Smith, 1992:1)，但其他全国性的定期调查的每次调查中也包含了相当多的重复问题。这些调查包括：全国选举研究(NES)，它自 1952 年起每两年都有可用的数据；全国健康访问调查(NHIS)，从 1957 年起每年搜集数据；人口现状调查(CPS)，这是一项基于轮换固定样本设计的、每月进行的、关于工作参与的调查。此外，偶然的民意调查，例如哥伦比亚广播公司与《纽约时报》的民意调查，也经常有重复的问题。其他有重复问题的重要数据还有消费者调查(密歇根调查研究中心)，它的数据包含一些能追溯至 1946 年的问题。这些数据都可以从密歇根大学的校际政治与社会研究协会(ICPSR)得到。有兴趣的读者可以参考协会最新的目录以及基尔科特和纳森的研究(Kiecolt & Nathan, 1985)，查看关于可用于研究社会变迁的全国数据的更多介绍。

重复调查与固定样本调查

为了研究变化，我们必须在不同的时间测量相同的事物。重复调查是对不同的人群进行测量。固定样本调查对相同的人群进行重复测量，所以固定样本调查是长时间追踪个人，而重复调查是随时间变化追踪不同的世代(同年出生的人群)。

哪个方法更好？对此没有一般性的结论，因为固定样本数据更适用于一些问题，而重复的截面数据更适用于另一些问题(Duncan & Kalton, 1987)。假设我们想知道在里根任期内，共和党人对民主党人的比例在适龄投票人口中有没有提高，用对该群体的重复调查就可以直接估计这个比例。例如，用全国选举

研究数据(Abramson, Aldrich & Rohde, 1994),我们估计这个比例从1980年的0.55(在党派人士中,35.4%是共和党人,64.6%是民主党人)提高到了1988年的0.78(43.8%是共和党人,56.3%是民主党人)。即使它仅意味着共和党候选人在20世纪80年代末比80年代初有更大的支持基础,但这也是个引人注目的提高。然而要注意,共和党人和民主党人的百分比变化趋势并不能告诉我们在里根任期内,有多少人的政党支持发生了变化。在极端情况下,假设没有一个人的政党支持发生过变化,那么我们观察到的共和党人对民主党人比例的全部提高就都来自世代更替(更老的、更倾向于民主党的世代被更年轻的、更倾向于共和党的世代代替)。在这种情况下,这个趋势夸大了政党支持的变化,实际上政党支持的变化为0。更可能的情况是,共和党人对民主党人比例的变化趋势低估了(而非高估了)政党支持变化的比例,因为人们会向两个方向改变,而这种互相抵消的变动没有被总和比例所反映。

由于重复调查不随时间跟踪个人,所以它们不能记录人们政党身份的"总变化",也就是说,重复调查不能记录人们政党支持的实际变化率。重复调查记录的是"净变化",即所有变化的净作用。

由于固定样本调查随时间跟踪个人,所以它们提供了对世代总变化和净变化的估计。用固定样本数据估计总变化的可靠程度,取决于变化速度相对于测量间隔来说有多快。大多数政党支持的变化都可能被每年的调查所观察到,因为人们一般不在一年内多次改变支持的政党。然而,其他政治现象则更多变。例如,人们可能经常改变对现任总统做得如何的看法;关于总统的受欢迎程度,每年测量一次则可能低估了个人的总变化。

就本书的目的而言,更重要的一点是,传统的固定样本设计不适用于估计人口总体的净变化。在传统的固定样本设计中,个人是在时间1被抽取并且跟踪,于是,在这种设计中,新的世代并没有随时间变化被加入样本。所以,我们不能估计人口总体的净变化。举例来说,我们不能使用传统的固定样本调查设计来判定在里根总统任期内,适龄投票人口中共和党人相对民主党人的比例是否增加了。

可见,无论哪种研究设计都不是万能的。重复调查不适合估计个人的总变化,而传统的固定样本调查不适合估计变化的总趋势。为了克服这些局限性,

调查方法论学者发展出包含固定样本设计和重复截面设计特征的调查设计。其中一种混合型的设计是“轮换固定样本调查”(Duncan & Kalton, 1987)。在轮换固定样本调查中,固定样本是轮换的——新的固定样本被加入,旧的固定样本被去除——以保证最新的样本反映了正在变化的总体。人口现况调查和收入与项目参与调查(SIPP)就是轮换固定样本调查的例子。另一种混合型的设计被称为“分批固定样本调查”(Duncan & Kalton, 1987),它在重复调查中加入固定样本,以解决用重复调查估计总变化时的困难。英国社会态度调查就是分批固定样本调查的例子。全国选举研究也可部分地当做分批固定样本调查,因为它在其中几次调查中包含了固定样本(Kiecolt & Nathan, 1985)。

基什建议使用分批固定样本设计(Kish, 1983、1986)。未来,分批固定样本设计和其他混合型设计可能会变得更普遍。然而,现有的调查数据通常不是重复截面数据(不包含固定样本),就是固定样本调查数据。有两本著作讨论了固定样本调查数据(Finkel, 1995; Markus, 1979)。在本书中,我们的关注点是分析重复调查,包括有着重复调查成分的分批固定样本设计及其他设计。

重复调查的分析设计

至少有3种基本的方法可用于分析重复调查。其中一种方法就是将各调查分开,分别进行分析。在这种方法下,分析重复调查与分析单独调查的方法是一样的,除了对多个调查重复同样的步骤。我想不出这么做的理由(除了在探索阶段对解释变量的作用是否随时间而变化进行初步分析)。如果研究目的是估计某个时点个体层次的关系,那么分析单独的调查就已足够。如果研究目的是分析一种关系是否改变,为了检验统计显著性,把调查“合并”起来而非单独分析它们则更容易(参看第6章)。即使变量的作用保持不变,把数据合并也更好,因为合并后可以对(恒定的)作用[1]估计得更准确。简单来说,如果研究者关注的是社会变迁,我就看不到有什么理由把重复调查分开分析,因此在本书中,我不建议这种方法。

第二种方法是用累积的截面数据来分析个人层次关系的大小和稳定性。

举例来说,研究者可以用这种方法来分析阶级和政党身份的关系是否随时间保持稳定。在通常情况下,这种累积数据方法不需要花功夫进行数据组织,因为被大量使用的重复调查都可以以这种累积数据的格式购买,如美国综合社会调查和全国选举研究。累积调查方法很灵活,只要研究者有检验个人层次或"微观"关系随时间变化的基础。第 5 章用一般的分解模型来呈现微观层面的变化如何导致总变化。第 6 章将描述和展示分析累积数据的简单模型,它们可用于检验微观层次的变化,也就是个人关系的变化。

第三种方法是用累积数据来分析总变化。本书的一个特点是,它强调对宏观层次变化的分析。第 3 章描述和展现了估计总变化的方法。第 4 章描述了如何将变化趋势分解为世代的净变化导致的变化以及世代更替导致的变化(也就是人口更替的作用)。因为世代分析框架是我们讨论总变化趋势的基础,所以第 2 章回顾了世代分析,讲述了分解世代效应、年龄作用和时期作用的方法。

关于术语

因为诸如"分解"这类术语在各个学科中有不同的含义,在本书中,为避免混淆,我用方程对概念进行正式表述,用例子来说明方程。"合并数据"可以作为说明术语的含义可能不明确的好例子。在本章中,我多次提到集合或"合并"调查。虽然这个术语在关于调查的文献中被使用,但必须注意,我们所说的集合调查而成的合并数据不能与计量经济学家所说的合并时间序列数据/截面数据相混淆(参照 Sayrs, 1989),后者即固定样本数据。也就是说,在计量经济学的文献中,会提及使用"合并数据"中的单位层次的特点来控制个体长久存在的特征,这些特征影响我们关注的因变量,但又难以测量(例如,一个人的成功欲望或者一个国家的地形,假如研究单位是国家)。显然,我们不能控制这些单位层次的特征,除非我们对给定的个体有多次测量,但我们不可能从合并的重复调查中得到它们。因此,"合并数据"有时指随时间跟踪相同的个人,有时指在不同时点得到的不同样本。为避免混淆,后文中我将用"累积数据"来指称集合起来的重复调查。

第 2 章 | 区分年龄、时期及世代效应

为尝试区分世代、年龄和历史时期的影响，世代研究提供了一个帮助我们研究社会变迁的模板。在本书中，我倚重这个模板，尤其是在讨论总变化的时候(第 3 章和第 4 章)。因为之后将大量引用世代研究的基本概念，所以在本章我将简单地对这些概念进行回顾。我从定义世代效应以及它与年龄作用、时期作用的区别说起，然后介绍解决区分世代、年龄和时期作用这一难题的不同的实证方法。熟悉世代研究的读者可以直接参看关于如何分解年龄、时期和世代效应的讨论。

年龄、时期和世代效应

社会科学家经常把世代研究作为他们研究变迁的起点。假设我们观察到一组成年人参与礼拜的比例有增长，这组成年人自青少年起我们就一直跟踪研究，我们如何解释这种增长？研究世代的人会立刻从两种过程考虑这个问题。这个增长可能来自与历史时期相关的一般事件或过程，也许在所研究的时期里出现了宗教兴趣的复苏(时期作用)。另一种可能是，这种增长来自与人们年长化或生命周期有关的更具体的过程(年龄作用)，随着年轻的成年人在一个社区中安顿下来、结婚、生育等等，他们更有可能参加礼拜，正如他们父母或祖父母经历过的那样。换句话说，我们研究的这个世代只是在重复以前几代人参加礼拜的生命轨迹。

当然，参与礼拜的比例也可能随时间的增加反映了年龄和时期作用的共同影响。也许生命周期作用和复苏作用共同使该世代参与礼拜的比例提高了，或

者生命周期的作用使参与礼拜的比例提高了，而世俗化（一种时期作用）使参与礼拜的比例降低了，因此我们观察到的提高反映了这两种互相抵消的作用的净作用（在此例中，生命周期的作用更强）。

请考虑另一种情况，我们在某一时点观察到不同年龄段的人的差别，而不是世代随时间变得年长而发生的变化。例如，我们发现20多岁的人比其他年龄段的人更不可能在总统选举中投票。世代研究者又会立刻想到两种不同的解释。第一种是年龄或者生命周期的解释：年轻人更难在一个社区中立足，他们更可能在上学，他们可能要照顾幼儿，诸如此类。如果年龄作用可以解释年轻人更少投票的行为，我们可以预期将来的世代会重复他们前辈随生命周期而变化的投票模式，即先是低投票率，到中年后该世代的投票率提高。另一种可能是，这个新的世代在他们的整个生命期都表现出低投票率，意味着这可能是世代效应，而不是年龄作用。在这种情况下，这一年轻世代的特别之处并不在于他们年轻，而是不同的世代在不同的时间出生。正如莱德所说，世代效应的出现是因为“任何世代的成员都只能参与生命的一小块，即他们在历史中占据的特殊位置”(Ryder，1965:844)。

简而言之，“世代效应”指世代由于相同的经历或反应而造成的不同，“年龄作用”指与年龄相关的影响所造成的变化，而“时期作用”指与历史时期相关的影响所造成的不同。世代效应来自世代特殊的经历和社会化，也来自世代对相同的历史事件的反应（例如，转折性事件可能对年轻人有更深刻的影响，比如肯尼迪总统被刺杀或伊朗人质事件）。[2]年龄作用可能基于“内在的”发展或成熟化的变化，或者与年长化有关的生理变化，也可能基于与年龄相关的生命周期过程（婚姻状况、为人父母）。时期作用指历史背景的一致作用，也就是时间或者历史状况一致地影响了所有的世代（参考Glenn，1977；Hagenaars，1990）。

识别的难题

研究者试图分解年龄、时期和世代效应时，总被以下方程所表现出的特性所困扰：

$$出生年份 = 测量年份 - 年龄 \quad [2.1]$$

方程 2.1 说明，出生年份或者世代与测量年份及年龄是线性关系。因此，如果我们把所有的变量——世代（出生年份）、时期（测量年份）和年龄——都作为自变量放入回归模型，我们就不能估计这个模型。也就是说，我们不能估计以下模型：

$$E(Y) = \beta_0 + \beta_P \times 时期 + \beta_A \times 年龄 + \beta_C \times 世代 \quad [2.2]$$

此处，$E(Y)$是 Y 的估计值，时期是测量年份，年龄用年表示，世代是出生年份（在此以及下文中，我用希腊字母来表示总体参数，所以 β 表示总体参数，而不是标准化的斜率）。

为了更好地理解为什么这里有估计上的问题，请考虑我们如何解释方程 2.2 中的 β_C。β_C 是控制了测量年份和年龄后，出生年份增长导致的 Y 的估计值的变化。然而，如果年龄和测量年份都被控制了，出生年份就不能变化，因为出生年份=测量年份一年龄。与之类似，时期的斜率 β_P 在控制了年龄和世代后也不能被估计，年龄的斜率 β_A 在控制了时期和世代后也不能被估计。这就是世代研究中的识别难题。

早先曾流传着两种看似有效然而错误的方法。用 Y 对所有组合进行回归——年龄与时期、年龄与世代、时期与世代——并没有什么意义。因为年龄与时期包含的信息跟年龄与世代或时期与世代所包含的信息一样，我们从各模型得到的决定系数（R^2）也相同，因此在这里，我们不能用“最佳拟合”作为决定最合适模型的标准。截面数据分析也不能剔除时期作用的可能性。运用截面数据分析，相当于我们用一个常数 k 代替了方程 2.1 中的测量年份，导致世代 $= k -$ 年龄。所以，我们仍然不能从年龄作用中区分出世代效应。因此，区分世代效应和年龄作用的难题不能通过运用截面数据分析来解决。

克服识别难题的策略

我们没有绝对可靠的分解年龄、时期和世代效应的技术手段（Wilmonth,

1990)。然而,我们有解决识别难题的可行策略。为了更具说服力,世代研究中经常使用以下的一种或多种策略(Firebaugh & Chen, 1995)。

事先控制系数

一种常用的策略是进行假定以帮助识别。这种策略的其中一种形式是,研究者假定这些作用中的一种——年龄、时期或世代——是 0,或至少小到可以安全地被忽略。举例来说,如果我们假定年龄作用可以安全地被忽略,那么我们就假定在方程 2.2 中,$\beta_A = 0$,然后只要用 Y 对时期和世代进行回归即可。

这个策略的问题在于,它的结果只有在用于识别的假定 $\beta_A = 0$ 成立的情况下才可靠。如果这个假定是错误的,那么我们对时期作用和世代效应的估计会被年龄作用所影响。把年龄 = 时期 − 世代代入方程 2.2,我们会很容易发现:

$$\begin{aligned} E(Y) &= \beta_0 + \beta_P \times \text{时期} + \beta_A \times \text{年龄} + \beta_C \times \text{世代(无法估计)} \\ &= \beta_0 + \beta_P \times \text{时期} + \beta_A \times (\text{时期} - \text{世代}) + \beta_C \times \text{世代} \\ &= \beta_0 + (\beta_P + \beta_A) \times \text{时期} + (\beta_C - \beta_A) \times \text{世代} \end{aligned} \quad [2.3]$$

换言之,如果我们用 Y 对时期和世代进行回归,得到估计值并不是一个问题。实际上,我们会得到对应于时期项和世代项的系数。问题在于,如何解释这两个系数。正如方程 2.3 所展示的,时期项的系数是时期作用与年龄作用的和($\beta_P + \beta_A$),而世代项的系数是时期作用与年龄作用的差($\beta_C - \beta_A$)。因此,方程 2.3 中的斜率只有在年龄作用实际是 0 的时候,才能得到时期和世代效应的无偏估计,否则,时期项的系数所反映的既包含年龄作用也包含时期作用,世代项的系数所反映的既包含年龄作用也包含世代效应。

我们也可以采用其他用于识别的假定。比如梅森、温斯伯勒和普尔等人展示了年龄、时期和世代效应在分类数据分析里可以通过限定邻近的世代(或年龄、或时期)的作用相同来估计(Mason, Mason, Winsborough & Poole, 1973)。

虽然一般来说,假定相邻类别的作用一致比简单地假定其中一种作用为 0 更审慎,但即使是相邻类别的作用一致的假定也有可能是错误的,这也会导致不可靠的结果(Glenn, 1976)。此外,由于年龄、时期和世代项之间的高度共线

性，我们需要大样本来获得可靠的估计。我们可以通过限制邻近类别有相同的作用来获得估计值，因为此时时期、世代和年龄不再具有完全共线性。虽然它们不再具有完全共线性，但是3个变量之间不免还是有高度共线性（在控制了第三个变量的情况下，任何两个变量之间的相关系数将接近于1.0）。因此，在小样本中，时期、世代和年龄作用的估计值容易不稳定。

使用辅助信息

当提到年龄—时期—世代识别难题的时候，我们指的是观测到的结果的模式在数学上可能被年龄、时期和世代效应中多于一种的组合所解释。然而，数学上的多种可能性，其大小在现实中并不完全相等。我们所观测到的结果有几种不同的解释，"辅助信息"帮助我们判断其中哪种解释的可能性相对更大(Glenn, 1977)。如果在3个预测变量（年龄、时期和世代）中的一个或更多变量之间存在非线性关系，它就给了我们这样的机会，因为通常对特定的一种非线性模式只有一种可能的解释（例如，工资与年龄之间的倒U型关系几乎肯定是年龄而不是世代对工资的作用）。

请看方程2.4，其中进入回归的是年龄和时期：

$$\begin{aligned} E(Y) &= \beta_0 + \beta_A \times \text{年龄} + \beta_C \times \text{世代} + \beta_P \times \text{时期（无法估计）} \\ &= \beta_0 + \beta_A \times \text{年龄} + \beta_C \times (\text{时期} - \text{年龄}) + \beta_P \times \text{时期} \qquad [2.4] \\ &= \beta_0 + (\beta_A - \beta_C) \times \text{年龄} + (\beta_P + \beta_C) \times \text{时期} \end{aligned}$$

假设$Y=$每年的礼拜参与（成年人每年参加礼拜的次数），假定我们用参与对时期和年龄进行回归后得到，$E(\text{参与}) = 5 + 0.5(\text{年龄})$。年龄的系数是正的，意味着年长的成年人更可能参加礼拜。时期的系数为0，显示各年龄人士参加礼拜随时间没有发生变化，即在给定的年龄，各相继的世代参加礼拜的比例是相同的。这个模式显示的是参加礼拜的次数会随着生命周期而增加，相继的世代会重复前辈参加礼拜的规律。数学上来说，这个模式可能来自几种年龄、时期和世代效应的组合。例如，$\beta_C = -0.5$且$\beta_P = 0.5$（负的世代效应完全被正的时期作用抵消，所以$\beta_P + \beta_C = 0$），且年龄没有作用（$\beta_A = 0$），因此$\beta_A - \beta_C = 0.5$。然

而，这样的组合很牵强。更可能的情况是，这种各世代参加礼拜的一致轨迹反映的是纯粹的年龄作用，即 $\beta_A = 0.5$(Firebaugh & Harley, 1991)。无论如何，重点是某些数学上的可能性比另一些更大，这个事实使我们分解年龄、时期和世代三者的作用成为可能(Glenn, 1976、1977)。

使用直接的测量

如果我们假定世代之间的差别来自各世代之间的相对大小，我们通常会使用直接测量的策略(Easterlin, 1980)。在这种情况下，世代(其大小)的明显特点很容易测量。在更多情况下，它的明显特点不那么清楚，且很难测量。请考虑莱德所讨论的世代效应，即“社会变迁”对年轻人的影响更大(Ryder, 1965: 861)。尽管我们都认为，年轻人实际上更具激情(Glenn, 1980; Mannheim, 1927/1952)，但鉴别和测量导致持续的世代差异的社会变迁是一项极为困难的工作。同样，在通常情况下，测量导致时期作用的社会变迁及导致年龄作用的、与年龄相关的影响，都是困难的工作。这种困难，部分是由于我们很难证明自己的测量是无遗漏的。举例来说，如果我们用关键的生命周期状况(婚姻状况、为人父母、工作状况等等)来测量年龄作用，总会有批评说，这些状况只是抓住了所谓“年龄作用”中一小部分基于年龄的影响。

简言之，我们没有“巧妙的方法”来区分年龄、时期和世代效应，但是有一些策略可用于解决识别难题。各策略的共同特点是需要仔细地将年龄、时期和世代效应的特质理论化。在“辅助信息”和直接测量策略中，理论的重要性显而易见。理论在“使用技术控制”的策略中同样重要，这种策略在模型中事先加入识别的约束条件。这些用于识别的约束条件必须基于审慎的理论，因为不正确的假定会导致不可靠的结果。如果有其他可能的识别假定，我们应该在不同的识别假定下，仔细比较年龄、时期和世代效应的估计值，以检查结果对不同的识别假定的敏感度。

第 3 章 | 总趋势

许多重复调查数据的搜集持续了几十年，可以用于发现社会趋势。其要点是应把趋势和偶然的变动区分开来。

平滑趋势

我们把 Y 从时间 1 到时间 2 的变化定义为 $Y_2 - Y_1$。对重复调查的分析有时会过多地解释这个差别，尤其是时间间隔较短的时候。从一项调查到另一项调查之间观察到的差异，有时被用于声称社会向一个或另一个方向变化的根据。人们应该小心此类声称。由于存在抽样误差和偶然的短期变动，仅靠两个时间点上的测量、凭直觉来推论社会趋势是轻率的。

许多重复调查已经累计搜集了 10 年或更多年或每两年一次的调查数据，这些调查为关于社会趋势的结论提供了更为坚实的基础。一项基本的分析原则是，对社会趋势的分析，数据多总比数据少好。通常我们不会仅仅分析头尾两个时间之间的差别，我们希望使用所有的调查。问题是，如何使调查数据之间的变动平滑，以判断是否存在潜在的趋势？

其中一种方法是使用移动平均值。我们设总的时间序列为 Y_1，Y_2，…，Y_T，其中 Y 是年份 1，2，…，T 的平均值。Y 在时间 t 的移动平均值由邻近的 Y_t 且包括 Y_t 的 Y 的加权平均值所决定。例如，有一个 5 年的移动平均值，Y 在时间 t 的移动平均值是$\{Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}\}$的加权平均值，其中权重可以依次为{1/9，2/9，3/9，2/9，1/9}（Kendall，1973：35）。加权平均值是有用的第一步，即便仅仅为了做一个平滑了的抽样波动图。平滑曲线图可以帮助判

断 Y 是否在给定的时间间隔内改变了变化方向。在这种情况下，用线性趋势来估计数据就会误入歧途。

如果用线性趋势来估计数据是合适的，我们就可以用回归来确定样本的趋势在统计上是不是显著。假设我们有 T 次重复的年度调查，合适的一步是把 Y_{it} 回归到调查年份上：

$$E(Y_{it}) = \beta_0 + \beta_1 \times 年_{it} \qquad [3.1]$$

$$其中，i = 1, 2, \cdots, I_t; \ t = 1, 2, \cdots, T \qquad [3.2]$$

$E(Y)$是 Y 的期望值，Y_{it}表示第 i^{th}个人在第 t^{th}次调查中的 Y 值，I_t 表示第 t^{th}次调查的样本量。此处，$E(Y_{it})$是条件均值（以年份的函数来表示的 Y 的期望值或均值）。一个非零的 β_1 意味着 Y 是年份的线性函数。斜率的符号表示 Y 的趋势是向上或向下的。在估计方程 3.1 的时候，将时间 t 为 1 时的年份（第一次调查的年份）作为 0 会很方便，由此 β_0 就是研究首年 Y 均值的预测值。

很多调查常常以较高的抽样概率抽取某些群体。在有些情况下，以较高概率抽样是有意为之的（保证样本中包含足够的少数群体成员）。在其他情况下，以较高概率抽样是抽样设计的副产品。例如，抽取户并且只访问户内的一个人，有些调查会以较高概率抽取独居的成年人。在这些情况下，可以通过加权来将之恢复为正确的人口百分比（Stephenson，1978）。样本加权的基本原则是用被抽中概率的倒数来加权（Kalton，1983）。假设我们的目标人口中有两个群体——M（多数群体）和 m（少数群体），分别构成人口的 90%和 10%。为了保证样本中有足够的少数群体成员，我们设定样本的构成为 0.20 的 m 和 0.80 的 M。为恢复正确的人口百分比，我们给 m 的成员加权 0.10/0.20 = 0.5，并且给 M 的成员加权 0.90/0.80 = 1.125。[3]

最后需要强调的是，使用方程 3.1 来检测趋势是假定了比较的对象在各调查中的测量是一致的（Smith，1993）。“机构影响”（Johnston，1981）构成了对这种一致性的威胁，所以对研究趋势来说，数据最好来自同一个调查机构。在不可能做到的情况下，有时候可以通过把两个调查机构的数据衔接起来以构成关于历史趋势的数据，只要这些调查在不同时点询问了同样的问题。这种衔接假

定，两个历史序列数据在时间上是重叠的。如果这些序列数据只是连接（没有重叠）而不是衔接，那么在两个序列数据的接点上即使有变化，也很难判断是真正的变化，还是由于数据来自不同的调查机构的影响。

即便数据来自同一个调查公司，也不能保证万无一失。有经验的调查研究者知道，有时候对调查问题措词的改变，即使看似无害，也可能导致回答上的显著差别（Rasinski，1988；Smith，1987）。此外，对调查中一个特定问题的回答，可能受到调查中其他问题的性质的影响，因此即使问题 Q 在各调查中保持一致，由于“上下文影响”，调查中其他问题的变化也可能导致对问题 Q 回答的变化（Smith，1988、1991）。即使措词和上下文在各调查中是一致的，关键词（例如“保守的”和“自由的”）的含义仍有可能随时间发生变化。在判断为什么在重复调查中出现了某种趋势时，人们首先应该问，这个趋势是不是仅仅反映了受访者理解调查问题的变化。

趋势中的组别差异：趋同还是趋异

整体趋势的方向可能掩盖了社会中主要群体的趋势。举例来说，调查显示，过去几十年中，反对黑人与白人通婚的比例下降了（Firebaugh & Davis，1988；Gallup & Newport，1991）。这个整体趋势可能掩盖了区域的或群体的差异。这种下降在南部地区是更快还是更慢？在男性和女性、黑人和白人中，这种反对比例的下降速度一样吗？这个趋势在白人女性和黑人女性中是不是出现了交叉，导致现在黑人女性比白人女性更可能反对种族通婚（Paset & Taylor，1991）？

在线性趋势的组别差异中，有以下几种可能的情况：一致趋势、平行趋势、趋同趋势、趋异趋势、交叉趋势。假设我们分别对南方人（S）和北方人（N）估计白人对种族通婚的反对趋势。各自的估计可以通过分开的回归方程得到：

$$E(Y_{it}^{S}) = \beta_0^{S} + \beta_1^{S} \times 年_{it} \qquad [3.3]$$

$$E(Y_{it}^{N}) = \beta_0^{N} + \beta_1^{N} \times 年_{it} \qquad [3.4]$$

其中 Y 是反对种族通婚的比例，如前所述，i 表示个人，t 表示调查。方程 3.3 估计的是南方人的线性趋势，方程 3.4 产生的是非南方人的线性趋势。如果我们设定第一次调查为年 = 0，那么两个方程的截距[4]分别为研究开始时南方人和北方人反对种族通婚的比例。为了简化注释，在此我假定测量间距为年，实际上测量间距可以是季度、月或者其他间距。在那些情况下，我们会相应地更改自变量，例如把方程 3.3 和方程 3.4 中的$年_{it}$改为$季_{it}$或$月_{it}$。

与分别估计对应于南方人和北方人的两个回归方程相比，通常更简便的方法是把两个方程合并为一个方程。方程 3.3 和方程 3.4 可以通过一个虚拟变量和一个交互项来合并：

$$E(Y_{it}) = \beta_0^* + \beta_1^* \times 年_{it} + \delta_0 S_{it} + \delta_1 (S \times 年)_{it} \qquad [3.5]$$

在这里，S 是一个虚拟变量，如果受访者是南方人则取值为 1。这种单独方程的方法提供的重要优点是，因为方程 3.5 中的 δ_0 和 δ_1 分别表示截距和斜率的差别，所以我们可以立刻判别出南方人和非南方人有没有开始状态的差异（δ_0）和趋势上的差异（δ_1）。

如果 δ_0 和 δ_1 同时为 0，那么南方人和非南方人的趋势是相同的（一样的截距和斜率）。如果 δ_0 不是 0 而 δ_1 是 0，那么他们的趋势是平行的。如果 δ_0 和 δ_1 都不为 0 且同号，那么他们的趋势是趋异的。如果 δ_0 和 δ_1 都不为 0 且异号，那么他们的趋势是趋同的（或者他们交叉了，这取决于南方人和北方人初始的差别以及趋同的速度）。

前述原则很容易推导出来。在方程 3.5 中代入 0 得到非南方人的估计，我们有：

$$E(Y_{it}) = \beta_0^* + \beta_1^* \times 年_{it} \qquad [3.6]$$

很清楚，这时 $\beta_0^* = \beta_0^N$（截距是非南方人的截距），且 $\beta_1^* = \beta_1^N$（斜率是非南方人的斜率）。给方程 3.5 代入 1 得到南方人的估计，我们得到：

$$E(Y_{it}) = (\beta_0^N + \delta_0) + (\beta_1^N + \delta_1) \times 年_{it} \qquad [3.7]$$

因为 $\beta_0^N = \beta_0^*$，且 $\beta_1^N = \beta_1^*$，因此，南方人的截距（β_0^S）和斜率（β_1^S）分别是 $\beta_0^N + \delta_0$ 和

$\beta_1^N + \delta_1$。所以，

$$\delta_0 = \beta_0^S - \beta_0^N \qquad [3.8]$$

并且

$$\delta_1 = \beta_1^S - \beta_1^N \qquad [3.9]$$

如果南方人的截距和斜率 δ 都不为 0 且为正，那么南方人在初始时间下对种族通婚的反对更大些（$\delta_0 > 0 \rightarrow \beta_0^S > \beta_0^N$），且 $\beta_1^S > \beta_1^N$，所以区域差异在增加。如果 δ 都为负，那么非南方人初始时间下的反对更大些（$\delta_0 < 0$），且因为 $\beta_1^N > \beta_1^S$，所以区域差异在增加。简言之，δ 有同样符号的时候趋势是趋异的，δ 有不同符号的时候则会出现趋同的趋势。

这些原则对"效果编码"同样适用，也就是说，在方程 3.5 中，S 对南方就编码为 1，对非南方则编码为 −1（在虚拟编码情况下，β_1^* 是参照组的趋势；在效果编码情况下，β_1^* 是南方人和非南方人趋势的均值）。把 +1 和 −1 分别代入方程 3.5 的南方人和北方人中，得到 $\beta_0^S = \beta_0 + \delta_0$，$\beta_1^S = \beta_1 + \delta_1$，$\beta_0^N = \beta_0 - \delta_0$ 和 $\beta_1^N = \beta_1 - \delta_1$。因此，

$$\delta_0 = (\beta_0^S - \beta_0^N)/2 \qquad [3.10]$$

且

$$\delta_1 = (\beta_1^S - \beta_1^N)/2 \text{（在效果编码的情况下）} \qquad [3.11]$$

关于自选择的注意事项

在某些情况下，由于成为某个群体成员的原因与我们感兴趣的变量相关，这也会造成群体间趋异的出现。举例来说，假设我们根据加入的教会类型把基督教徒分成"保守的"和"自由的"，然后我们发现这两组人在公立学校的祷告问题上越来越两极化。在没有固定样本数据的情况下，我们难以解释这种两极化，是因为保守的基督教徒变得比自由的基督教徒更支持祷告吗？或者因果关系的方向是相反的，即对于祷告的态度影响了人们加入哪个教会？换言之，当公立学校的祷告问题变成一项更引人关注的公共事件时，也许它也变成了选择

教会的一项重要标准，且这可以解释我们观察到的两极化，因为如果自由教会中的成员支持在公立学校中祷告，就可能转而加入保守教会，而如果保守教会中的成员反对在公立学校中祷告，就可能转而加入自由教会。

这个问题的一种变形，可以从南方人和非南方人在反对黑人歧视方面的差异缩小的发现中看到（Firebaugh & Davis，1988）。可能是迁移的结果，区域差异的缩小可能是"虚假的"，因为迁移提高了来自非南方地区的南方人的比例。为了排除这个可能，菲尔鲍和戴维斯在包含移民和不包含移民的情况下研究了区域趋势（并且得到了相似的结果）（Firebaugh & Davis，1988）。

此处我们要强调的是，在使用重复调查数据来检验趋同和趋异趋势的时候，要注意自选择和迁移问题。这个问题在个人能够选择加入或离开某个群体——宗教群体、政党、地域等等（但不包括先赋的群体，如年龄群或性别）——的时候就容易出现，因为加入和离开这些群体的因素可能与因变量相关。

趋异模型的实证案例：检验年龄两极化的假设

在当前美国，我们关注退休人员和劳动者之间的断裂。记者们猜测，两个群体间相对幸福程度的改变与长者福利成本的提高，共同造成了年轻人与长者间的潜在冲突，这体现在社会福利、医疗卫生、教育以及其他在利益上可能趋异的领域中。例如，多家报社的专栏作家理查德·里夫斯（Richard Reeves）警告说："美国出现了长者的要求与全社会的需要之间的对立。"（Rosenbaum & Button，1992:385）朗曼声称："婴儿潮时期出生的人正在将他们收入中前所未有的部分来支持现在已经退休的年长的一代人。"他也警告说："很可能的结果是发生一场年轻人与长者的战争，除非许多根本性的趋势发生了迅速转变。"（Longman，1987:2）相反，老年学家一般认为："老年人和年轻人在大多数事情上没有什么区别。"（Day，1990:47）

虽然发生一场代际战争是牵强的，但在美国，与年龄相关的区别也许正在深化。单公式的交互模型（方程 3.5）在此情况下适用。如果这种区别在深化，我们预期会得到相同符号的 δ，这意味着趋异的趋势。

我选择两个因变量来检验年龄两极化的假设:对教育支出的支持和对社会福利支出的支持。它们是从美国综合社会调查的一系列问题中选出来的,其介绍如下:

> 我们的国家面对很多问题,它们中没有一个可以很容易或花费很少的钱而解决。我会提及其中的一些问题,希望你告诉我,你认为我们是否在解决它们的时候花费了太多、太少或适量的钱?我们是否花费了太多、太少或适量的钱……在提高国家的教育系统和社会福利上?

对教育(支出)的支持最早于1973年加入问卷,对社会福利(支出)的支持于1984年加入问卷。我选择教育是因为在一项截面研究中(1988年全国选举研究),文诺夫斯基发现了"颇强"的年龄差别,其中长者对教育支出的支持度最低(Vinovskis, 1993:62)。我选择社会福利是因为它在代际平等的文献中的中心地位(Donza, Duncan, Corcoran & Groskind, 1988; Kotlikoff, 1992)。对这两个因变量,我对其回答进行了两种区分——太多对适量或太少以及太少对适量或太多——并使用 logistic 回归。

表3.1和表3.2汇报了结果。退休人员确实显著地更不可能支持教育支出(表3.1中的δ_0),他们更可能说我们在上面花费了太多,而更不可能说我们花费得太少。然而,与年龄两极化假设不符的是,这个差别(δ_1)在过去20多年中并没有变得更明显。在任何一种编码方法下,初始状态的差别和趋势的差别都有不同的符号,这都反映了趋同而不是趋异。无论如何,尽管样本相对较大,反映趋势差别的系数(δ_1)都没有达到统计显著性。我们认为,退休人员和劳动者的趋势是平行的。总的来说,退休人员更不支持教育支出,但是没有证据表明长者和年轻人在该问题上的区别越来越大。

对于社会福利支出的情况(表3.2),结论取决于变量如何区分。与代际冲突的文献相一致,退休人员不可能说我们在社会福利上支出太多,但是他们也更不可能说我们在社会福利上花费太少。退休人员显著地更可能回答,在社会福利上的花费是"适量的",而其他人更可能选择"太多"或"太少"。对于劳动者或退休人员谁更可能支持社会福利开支这个问题,取决于我们研究的是分布的哪一端。

表 3.1 对教育支出的支持的趋势分析(1973—1993 年:logic 系数)

模型	因变量的编码[a]	
	1 = 支出太多	1 = 支出太少
模型 1:退休人员对其他人[b]($N = 19012$)		
初始差别:退休人员减其他人[c](方程 3.5 中的 $\hat{\delta}_0$)	0.978**	−0.583**
退休人员的趋势($\hat{\beta}_1^* + \hat{\delta}_1$)	−0.075**	0.050**
趋势的差别:退休人员减其他人[d]($\hat{\delta}_1$)	−0.019	0.001
模型 2:退休人员对有薪水的劳动者[b]($N = 13992$)		
初始差别:退休人员减劳动者[c]	0.947**	−0.643**
退休人员的趋势	−0.075**	−0.050**
趋势的差别:退休人员减劳动者[d]	−0.013	0.002

注:a. 在第一栏的系数中,“适量”与“花费太少”归为一类,在第二栏中,它与“花费太多”归为一类。第一栏:1 = 花费太多,0 = 适量或花费太少;第二栏:1 = 花费太少,0 = 适度或花费太多。
b. “其他人”包括处理家务者和学生,“有薪水的劳动者”不包括前述群体。
c. 截距的差别,其中第一年(此处是 1973 年)编码为 0。
d. 趋势的差别(年的斜率)。
* 表示 $p < 0.05$; ** 表示 $p < 0.001$。
资料来源:美国综合社会调查(Davis & Smith, 1994),不包含以较高概率抽取的黑人样本。

表 3.2 对社会福利支出的支持的趋势分析(1984—1993 年:logic 系数)

模型	因变量的编码[a]	
	1 = 支出太多	1 = 支出太少
模型 1:退休人员对其他人[b]($N = 12262$)		
初始差别:退休人员减其他人[c]	−0.879**	−0.481**
退休人员的趋势	0.042	−0.008
趋势的差别:退休人员减其他人[d]	0.078	0.022
模型 2:退休人员对有薪水的劳动者[b]($N = 9666$)		
初始差别:退休人员减劳动者[c]	−1.029**	−0.489**
退休人员的趋势	0.042	−0.008
趋势的差别:退休人员减劳动者[d]	0.080	0.029

注:a. 在第一栏的系数中,“适量”与“花费太少”归为一类,在第二栏中,它与“花费太多”归为一类。第一栏:1 = 花费太多,0 = 适量或花费太少;第二栏:1 = 花费太少,0 = 适度或花费太多。
b. “其他人”包括处理家务者和学生,“有薪水的劳动者”不包括前述群体。
c. 截距的差别,其中第一年(此处是 1984 年)编码为 0。
d. 趋势的差别(年的斜率)。
* 表示 $p < 0.05$; ** 表示 $p < 0.001$。
资料来源:美国综合社会调查,不包含以较高概率抽取的黑人样本。

同样，没有证据显示有趋异的趋势，至少对有数据的这 10 年（1984—1994 年）来说是这样。交互项 δ_1 没有达到统计显著性。尽管退休人员和劳动者之间存在差别，但这种差别在过去 10 多年中保持了稳定。

除了检验年龄两极化假设，趋同/趋异模型对当代社会问题还有其他明显的应用。例如，最近我们听到"男女差别"在支持共和党、环保态度、机会均等行动的态度等方面的表现。男女之间在态度和行为上是否有差别？如果有，近几十年来有没有扩大？

福克斯和菲尔鲍使用趋势趋同/趋异模型和综合社会调查来回答不同性别在对科学的信心上的差别（Fox & Firebaugh, 1992）。他们发现了趋异的趋势：女性比男性对科学更没有信心，且这种差别在过去几十年中扩大了。用以上的方程来说，初始差异的系数 δ_0 和趋势差异的系数 δ_1 有相同的符号，且在研究中都达到了统计显著性。因此，他们的研究指出了性别两极化的可能性，至少在美国人对科学的信心方面来说是如此。

第 4 章 | 分解总趋势

如果我们发现，在选民中民主党人和共和党人的相对比例发生了变化，我们就可以推论，要么是个人改变了政党支持，要么是选民的总体发生了变化(假定是因为年长的世代去世，并且被年轻的世代所代替)，或者两者都有。换言之，总趋势的可能的来源是个人和总体的改变所造成的净变化。作为研究社会变迁的第一步，从成员身份的变化中区分出个人改变造成的变化通常是有用的。在本章中，我会描述和演示两种“近似分解”方法，也就是把一个趋势分解为它们可能的来源。我会以介绍在什么情况下，社会变迁的速度会快于社会中一般成年人的变化来总结本章。

世代内的变化对总变化

重复调查可用于研究社会变迁。例如，白人受访者对反对黑人和白人通婚的法律的支持从 1974 年美国综合社会调查显示的 35%下降到 1994 年美国综合社会调查的 16%。为了研究产生这种变化的可能的原因，追踪世代内的百分比变化是有用的。如果世代内的变化与总变化保持一致，我们就可以推论说，总变化从个人变化的净效应而来。相反，如果在世代内反对种族通婚的比例没有随时间而改变，我们可以推论说，总变化来自人口更替的变化，而不是个人变化的净效应。更常见的情况是，总变化来自上述两种变化，此时的重点是判别它们各自的相对贡献。

表 4.1 展示了把数据排列起来以比较世代内的变化与总变化的有用方法。行表示世代，因此从左向右阅读每行，我们可以发现，每个世代内白人对反

对种族通婚的法律的支持率变化。世代内的变化可以与总百分比的变化相比较：1974 年到 1984 年的变化是－7.3%，1984 年到 1994 年的变化是－11.4%（表 4.1）。

表 4.1 中的总百分比是世代百分比的加权平均值。例如，1974 年，34.6%的受访者对反对种族通婚的法律表示支持。这个百分比是第一列内的世代百分比的加权平均值：[15.0 × (307/1243)] + [24.3 × (235/1243)] + [32.8 × (201/1243)] + [45.2 × (217/1243)] + [50.3 × (159/1243)] + [71.4 × (98/1243)]+[61.5×(26/1243)] = 34.6（在四舍五入误差下）。因为总百分比是世代百分比的加权平均值，所以总百分比的变化是由世代百分比的变化或者世代的相对大小（权重）的变化而来，或者两者均有。简言之，社会变迁可以看做世代内的变化和世代相对大小的变化的函数。

表 4.1　美国成年人[a] 反对种族通婚态度的世代内的变化和总变化（百分比）：以 10 年为间隔（1974—1994 年）

世　代	1974		1984		1994		变　化	
	百分比	频数	百分比	频数	百分比	频数	1974—1984	1984—1994
1965—1974	—	—	—	—	8.5	258	—	—
1955—1964	—	—	14.9	302	8.9	404	—	−6.0*
1945—1954	15.0	307	13.9	267	10.1	346	−1.1	−3.8
1935—1944	24.3	235	27.4	179	16.5	218	+3.1	−10.9*
1925—1934	32.8	201	38.1	139	29.6	159	+5.3	−8.5
1915—1924	45.2	217	42.7	157	37.1	140	−2.5	−5.6
1905—1914	50.3	159	57.0	93	41.7	60	+6.7	−15.3
1895—1904	71.4	98	52.8	36	—	—	−18.6	—
1885—1894	61.5	26	—	—	—	—	—	—
所有世代	34.6	1243	27.3	1173	15.8	1585	−7.3*	−11.5*
世代内变化的均值（以世代大小加权）							+0.4	−7.1*

注：a. 白人，年龄等于或大于 20 岁。
* 表示 $p < 0.05$。

需要强调的是，表 4.1 的编排有特殊的意义。之所以这样，是因为列均值的改变（包括百分比，因为百分比是取值为 0 到 100 的二分变量的均值）总是可以表达为行内均值变化的加权平均值。因为这总是成立，那么表 4.1 的编排有什

么特别？表 4.1 的编排有重要意义，因为通过选择世代为行变量，它区分了基于个人的和基于总体改变的社会变迁。因为个人内嵌于世代中，所以行以内的改变是个人的改变。当然，我们不知道行与行之间的变化来自年龄作用还是时期作用（对照上述的方程 2.3；Firebaugh，1990；Rodgers，1990），但这并不是此处关注的问题。问题是，社会变迁中有多少来自观念的改变，有多少来自年长的成年人被更替为年轻的成年人？詹姆斯·戴维斯这么说：

> 毫无疑问，在同一行中，我们讨论的是同一（群）人……在每一列中，我们讨论的是不同的总体。由此，这个编排确实把变迁区分为“改变”（在特定行中的人改变他们的观点）和“更替”（列变量的构成的变化）。（James Davis，1992:274）

诺尔波特（Norpoth，1987）认识到世代—时期的数据排列的重要意义，他用这种数据排列回答了本章开头所提出的改变政党倾向的问题。在他的表 4（Norpoth，1987:386）中，诺尔波特运用世代—时期排列，主张由于世代更替，转向共和党的政党倾向改变“正在发生并且持续进行着”。用他的话来说，“自 1980 年来，年轻人中共和党支持者的历史性增长……预示了通过代际更替发生的政党倾向改变”（Norpoth，1987:376）。然而诺尔波特由此止步，实际上，他没有把共和党人变化的百分比分解为世代更替和观念改变两部分。在这方面，诺尔波特的研究并不特别。虽然社会科学家经常提到世代或代际更替的影响，但他们很少去估计这种影响的大小，原因可能是，他们知道估计代际更替作用的人不多。

在本章余下的部分，我会介绍两种方法——一种基于回归，另一种基于代数——来把社会变迁分解为世代更替和世代内的改变两部分。对种族通婚的态度改变是一个方便的起点。通过把数据排列成表 4.1 的形式，我们可以立刻看到对种族通婚的反对从 1974 年到 1984 年之间的下降只来自总体改变；相反，从 1984 年到 1994 年，这些改变大多来自态度改变（下文详述）。从 1974 年到 1984 年，世代内改变的平均值接近 0。一些世代的均值改变为正，一些为负，但

这些改变都没有达到统计显著性，因此，7.3%的总下降来自总体的改变，即年长的、更带偏见的世代被年轻的、较少偏见的世代所代替。

菲尔鲍和戴维斯研究了 1972 年至 1984 年间人们对种族通婚态度的改变，得出了一样的结论(Firebaugh & Davis, 1988)。他们不是把数据编排成表 4.1 的形式，而是用回归来分解总趋势。下一节将介绍他们使用的方法。

线性分解

在本节中，我将介绍“线性分解”(Firebaugh, 1989)的方法。线性分解假定，世代内的改变都是线性且叠加的。在下一节中，我还将介绍一种代数分解方法，它用于线性—叠加的假设不恰当的情况下。按照以往的经验，线性分解通常比代数分解更容易，但这两种方法在变化是单调的情况下，所得出的结论是相似的。

线性分解和代数分解都把总的社会变迁分解为源于世代更替的部分和个人改变的部分。虽然线性回归运用了世代—时期的设计，但它并不像表 4.1 那样从把某些相邻出生年的人合并成一组开始，因为线性分解把世代看做连续的(出生年)。线性分解的第一步不是标准的世代表，而是线性和叠加的回归模型。

因为世代表(例如表 4.1)是世代分析的常规方法，所以我们选择何处作为起点来分解变迁很重要。在标准的世代表中，行和列的类别有相同的宽度(例如，都以 10 年为间隔)，所以可以在表的对角线中跟踪同龄群或世代(在世代—时期设计中，同龄群在表的对角线上移动)。因为年龄要根据列的间隔来合并成群，所以信息便浪费了。然而，把年龄按对角线排列来分解变迁是不必要的。分解的目的是区分由个人改变而导致的变化和由世代构成改变而导致的变化。为了做到这点，我们只需要把数据排列成我们能够随时间跟踪世代的形式。为了随时间跟踪世代，我们可以采用一种简单的数据排列，以出生年为行变量，时期(测量年份)为列变量。因此，对于社会变迁的分解，我们没有必要把出生年合并为几个大类，如“大萧条世代”，“第二次世界大战世代”，等等。如果研究者

决定合并成群，以理论来指导合并可以避免世代在表中与列的时间间隔相匹配的苛刻要求。

进行线性分解

线性分解包括两步。第一步是用回归来估计世代内 Y 的年变化。因为我们假定世代内的斜率是线性的和平行的（叠加的），所以我们可以用以下的回归方程来估计世代内的年变化：

$$Y_{it} = b_0 + b_1 \times \text{年}_{it} + b_2 \times \text{世代}_{it} + e_{it} \qquad [4.1]$$

其中，Y_{it}是在第 t 次调查中第 i 个受访者的 Y 值，b_0 是估计的截距，b_1 是估计的世代内的斜率，b_2 是估计的世代间的斜率，年_{it}是第 t 次调查中第 i 个受访者的调查年，世代_{it}是第 t 次调查中第 i 个受访者的出生年。请注意，累积数据被用于估计方程 4.1。因为世代在方程中，所以 b_1 估计的是控制了世代内的斜率之后的世代变化。世代的系数 b_2 是世代间的斜率，即相邻世代间的平均差别。[5]线性分解假定世代间和世代内的斜率是线性的。

线性分解的第二步是用方程 4.1 中的斜率来估计世代内的改变和世代更替对总的社会变迁的贡献。因为 b_1 估计每个时间单位（这里为年）里世代内的改变，为了估计世代内改变对社会变迁的总贡献，我们把 b_1 乘以从第一个调查到最后一个调查的间隔年数（或季数、月数，取决于时间单位）：

$$\text{估计的世代内变化的贡献} = b_1(YR_T - YR_1) \qquad [4.2]$$

其中，YR_T 是最后一次调查的年份，YR_1 是第一次调查的年份。同样，为估计世代更替的贡献，我们把 b_2 乘以从调查 1 到调查 T 内出生年的均值的改变：

$$\text{估计的世代内变化的贡献} = b_2(C_T - C_1) \qquad [4.3]$$

其中，C_T 是最后一次调查样本中的平均出生年，C_1 是第一次调查样本中的平均出生年。这两部分加起来通常不完全等于总变化，但差别不应该很大，如果差别很大，我们的线性—叠加假设就有问题，意味着我们应该用另一种分解方法。

本节证明，在各部分与总社会变迁是线性的和叠加的关系的时候，将方程4.2和方程4.3所给出的各部分相加确实等于总的社会变迁。考虑把Y回归到年和世代的总体模型：

$$Y_{it} = \beta_0 + \beta_1 \times 年_{it} + \beta_2 \times 世代_{it} + \varepsilon_{it} \qquad [4.4]$$

在此，希腊字母代表总体的参数。同前，β_1是世代内Y的年变化，β_2是世代间的斜率。世代间的斜率反映世代在不同的时点上Y的不同。如果所有世代在任何时点有相同的均值，则β_2为0。

Y的均值(以$\bar{Y}$表示)是Y的期望值。在通常的假定$E(\varepsilon)=0$下，由方程4.4可得最后一次调查(调查T)的Y的均值是：

$$\begin{aligned}\bar{Y} &= E(Y_{iT}) \\ &= E(\beta_0 + \beta_1 \times 年_{iT} + \beta_2 \times 世代_{iT} + \varepsilon_{iT}) \\ &= \beta_0 + \beta_1 \times E(年_{iT}) + \beta_2 \times E(世代_{iT}) + E(\varepsilon_{iT}) \\ &= \beta_0 + \beta_1 YR_T + \beta_2 C_T\end{aligned} \qquad [4.5]$$

(重复以上)在此，C_T是最后一次调查的受访者的平均出生年。同理，第一次调查的Y的平均值是：

$$\bar{Y}_1 = \beta_0 + \beta_1 YR_1 + \beta_2 C_1 \qquad [4.6]$$

从第一次调查到最后一次调查的总变化是：

$$\bar{Y}_T - \bar{Y}_1 = \beta_1(YR_T - YR_1) + \beta_2(C_T - C_1) \qquad [4.7]$$

在线性—叠加情况下，简单的回归方法就可以把总变化$\bar{Y}_T - \bar{Y}_1$分解为个人变化的部分和更替的部分。

实证案例：歧视黑人的趋势

菲尔鲍和戴维斯用线性分解法研究了白人对黑人歧视的趋势变化(Firebaugh & Davis, 1988)。首先，他们根据1972年、1976年、1980年和1984年美

国综合社会调查的 4 个问题设计了一个歧视指数。这个歧视指数在 1972 年到 1984 年间下降了 1.22。把歧视指数回归到年和世代上，得到 $b_1 = -0.0457$ 和 $b_2 = -0.0508$ (Firebaugh & Davis, 1988;表 1)。把它们代入方程 4.2，估计出的世代内的变化对总下降的贡献是 $-0.0457 \times (1984-1972) = -0.55$。

在 1972 年样本中，白人的平均出生年是 1927.2，在 1984 年样本中是 1939.8。因此，由方程 4.3，世代更替对总变化的贡献估计是 $-0.0508 \times (1939.8 - 1927.2) = -0.64$。把两部分加起来得到 -1.19，接近于观察到的变化 -1.22。我们得出的结论是，从 1972 年到 1984 年，美国传统的对黑人的歧视下降了，且该下降的一半或一半以上源自年长的、更带偏见的世代被年轻的、较少偏见的世代所代替。

代数分解

1955 年，埃弗兰・北川(Evelyn Kitagawa)证明了两个人口群间比率(例如生育率)的差异是以下差异的函数:(1)分年龄比率的差异(对分年龄比率的进一步分解，参见 Das Gupta, 1993;例子参见 Smith, Morgan & Koropeckyj-Cox, 1996);(2)年龄构成的差异;(3)差异(1)与差异(2)的乘积。此处，我们要分解的不是在同一时点下不同人口群或者国家之间的差别，而是想分解同一个国家在不同时点下的差异，我们关注的是世代的构成变化，而不是年龄的构成变化。然而，逻辑上是相似的，只要进行一些修改，北川的分解方程(Kitagawa, 1955)就可用于我们希望进行的这种分解。

首先，我们把概念正式化，即一个群体的均值等于各子群在总体中的比重加权后的均值的和:

$$\mu = \sum_j p_j \mu_j \qquad [4.8]$$

此处，$\sum$ 表示加总，μ_j 是第 j 子群的 Y 的均值，p_j 是第 j 子群占总体的比重(p_j 加总后为 1.0)。由此得到群体 1 和群体 2 均值之差($\Delta\mu$)是:

$$\begin{aligned}\Delta\mu &= \mu_2 - \mu_1 \\ &= \sum\nolimits_j p_{j2}\mu_{j2} - \sum\nolimits_j p_{j1}\mu_{j1}\end{aligned} \qquad [4.9]$$

此处为区分群体 1 和群体 2,我们用了下脚标。对方程 4.9 进行代数处理可得(Kitagawa, 1955):

$$\Delta\mu = \sum\nolimits_j p_{j1}\Delta\mu_j + \sum\nolimits_j \mu_{j1}\Delta p_j + \sum\nolimits_j \Delta\mu_j\Delta p_j \qquad [4.10]$$

此处,$\Delta\mu_j$ 等于 $\mu_{j2}-\mu_{j1}$,即两个群体(例如国家)在子群 j 的 Y 的均值上的差,Δp_j 等于 $p_{j2}-p_{j1}$,即各群体中子群 j 占总体比重的差。方程 4.10 上的 3 个相加的项分别对应"比率"成分、构成成分和乘积成分。

方程 4.10 是对比率进行分解的最根本的方程。为了把比率分解方程应用到分析社会变迁中,我们把它修改如下:

用 Δ 表示随时间的变化;
用下脚标 1 和 2 表示调查 1 和调查 2,而不是群体 1 和群体 2;
用下脚标 j 表示第 j 世代。

最后一项修改是重点。通过用 j 来表示世代,方程 4.10 中的第一项变成了 $\Delta\mu_j$ 的加权和,即世代内 Y 的变化。这个加权和反映了个人的净变化对总变迁的贡献,因为个人的变化就是世代内的变化。如果总体内的各群体比重不变,对所有世代来说,Δp_j 都等于 0,那么方程 4.10 中的第二项和第三项都为 0。在这种情况下,所有变化都来自个人的变化,方程 4.10 的第一项理所应当捕捉了所有的变化。

方程 4.10 中的第二项——Δp_j 的加权和,即世代在总体中比重的变化——用于反映世代更替。在个人不发生变化的情况下,$\Delta\mu_j$ 等于 0,第一项和第三项则都为 0。在这种情况下,所有变化都来自世代更替,方程 4.10 的第二项理所应当捕捉了所有的变化。

方程 4.10 的第三项反映的是 $\Delta\mu$ 中不单纯源于个人变化或世代更替的部分。这个共同作用通常相对较小,因为它是两项变化的乘积。达斯·古普塔建议把它平均地分布在第一项和第二项中(Das Gupta, 1987)。这个方法产生了

一个包含两个组成成分的方程：

$$\Delta\mu = \sum_{j}[(p_{j1} + p_{j2})/2]\Delta\mu_{j} + \sum_{j}[(\mu_{j1} + \mu_{j2})/2]\Delta p_{j} \qquad [4.11]$$

三成分(方程 4.10)与两成分(方程 4.11)分解法的差别只在于给 $\Delta\mu_j$ 和 Δp_j 的权重。在三成分的情况下，第 j 世代的世代内改变由该世代在总体中的初始比重(它在第一次调查中占总体的比重)来加权，在两成分的情况下则由世代占总体的平均比重 $(p_{j1}+p_{j2})/2$ 来加权。同理，在三成分的情况下，第 j 世代占总体比重的改变由该世代的初始均值加权，在两成分的情况下则由该世代的平均均值来加权(两成分的分解法的一个变形将在后面分解对性别角色的态度改变的例子中提到)。

需要重点强调的是，在重复调查中，如果没有固定样本成分，我们就只能随时间跟踪世代而不是个人。如果世代内的死亡率与 Y 相关，那么我们观察到的世代内的变化就是个人变化和源自死亡率的变化的混合物。用世代内的变化来估计个人变化，我们必须假定 Y 与世代内的死亡率不相关。这个假设有时候是有问题的，尤其在对群体的趋势进行分解而群体又有不同的分年龄的死亡率时。例如，在分解人口中共和党人的百分比变化趋势时，我们必须假定共和党人和非共和党人有着相同的分年龄死亡率。如果在重复调查中有可靠的固定样本成分，研究者就能检验这些假设。这是在重复调查中加入固定样本成分的原因之一。

实证案例：再论歧视黑人的趋势

我们再把 1972 年至 1984 年的歧视黑人指数的总变化进行分解，这次使用代数分解法(SPSS 程序参看 Firebaugh，1992)。我们用两成分的分解法，使之可以与线性分解法的结果相比较。估计值如下：个人变化 = −0.51，世代更替 = −0.71。对应的线性估计值是−0.55 和−0.64。因此，两种方法产生了本质上相似的结论：歧视黑人的下降同时源自个人变化和世代更替。然而，代数分解法更多地把这种改变归因于世代更替。[6]

为什么总变化快于个人变化?

我们现在有更好的机会来理解为什么成年人经常抱怨变化太快了。当世代更替加强了个人变化且总变化快于平均的个人变化时,从典型的成年人的角度看,社会变迁实际上"太快了"。

斜率符号相同的规则

在线性—叠加作用的标准假设下,当世代内和世代间的斜率有相同的符号时,总变化会快于个人变化。再考虑方程 4.7,注意 $YR_T - YR_1$ 总是大于 0,且除非年轻人的死亡率大于年长者的死亡率,这种情况不太可能出现,否则 $C_T - C_1$ 大于 0。由此,只有在 β_1 和 β_2 同号时,个人变化和世代更替对变迁的贡献是同向的。由此得到斜率符号相同的规则:假定年长成年人的死亡率高于年轻人,当世代内和世代间的斜率的符号相同的时候,社会变迁就快于一般成年人的变化。

因为 β_1 和 β_2 是年龄作用、时期作用和世代效应的函数,所以这个相同符号规则可以与大多数社会科学家熟悉的概念相关联。考虑世代间的斜率 β_2。为什么成年人世代有差异?一种流行的观点认为,由于在生命期内,他们被暴露在不同的历史状况下,所以出生于不同年代的成年人有不同的信仰、态度、价值等等。此外,如果态度在年轻人中更易变,在年长人中更坚定(Gleen, 1980),那么相同的历史事件对年轻人作用更大,也可能造成世代之间的差别。

这些观点在两篇经典文章中被强调:卡尔·曼海姆(Karl Mannheim)的《关于世代的问题》(*The Problem of Generation*)和诺曼·莱德(Norman Ryder)的《世代作为研究社会变迁的概念》(*The Cohort as a Concept in the Study of Social Change*)。它们都假定,刚进入成年人世界的年轻人比已经在那个世界中的人更容易改变,因为年龄更大的人更坚持他们年轻时获得的看法。曼海姆写道:"早年的印象易于形成为对世界的自然看法。"(Mannheim, 1927/1952:298)因此,"所有具体经验都从它与初期的经验层的联系中获得它的特有表象和形

式”，曼海姆认为这些经验在大约 17 岁时获得。莱德写道：“每一个新的世代都与当代的社会遗产发生新的联系，且这种相遇的烙印会保留一生。”(Ryder，1965:844)简言之，“世代在生命的早年发展出特有的世界观定义，而这些看法似乎将持续整个成年期”(Lesthaege & Surkyn，1988:40)。

当年长的世代被后来的世代所代替，世代的“特有的世界观定义”会引起社会变迁。这就是孔德在书中说，“我们社会过程依赖于人的死亡”的含义(Comte，1839/1974，第 6 卷，第 6 章：518)。莱德进一步主张，社会变迁一旦开始，就可能因为一系列内在的动力而持续(Ryder，1965)。这个动力来自“人口新陈代谢”的双过程——持续的年长世代的逝去和年轻世代的加入——和世代差异。如果“世代的意义被植入”，变化将预示进一步的变化(Ryder，1965:861)。为了植入世代的意义，“社会世界的变迁必须用不同方式改变不同年龄的人”，且“这些变迁的作用(必须)持续”(Ryder，1965:861)。因为社会变迁对年轻人有更大的作用，所以改变将延绵到未来，因为年长的、受影响更少的世代逐渐逝去，而年轻的、受影响更大的世代加入了成人世界。

简言之，社会变迁既有即时的、可观察到的影响(因为它一定会影响现有人口的部分成员)，也有延迟影响。在一个没有世代延续的假设世界里，社会变迁的作用将只是即时的，或者如果是延迟的，就只是延迟的，因为它需要时间来让影响在人口中传播。在现实世界中，世代延续“放大”了社会变迁的作用，这种放大不是通过传播，而是通过使人口构成从受影响更小的世代改变为受影响更大的世代。

因此，在莱德的模型中，总变化快于一般人的变化是由于世代效应与时期作用互补。由同号原则，我们可以更正式地推导出在什么样的年龄—时期—世代的情况下，总变化快于个人变化。

斜率同号原则和年龄—时期—世代效应

世代内和世代间的斜率的符号取决于年龄、时期和世代效应的符号和相对大小。这在包含所有 3 种作用的线性—叠加模型中最容易看明白：

$$Y_{it}=\beta_0+\beta_P\times 时期_{it}+\beta_C\times 世代_{it}+\beta_A\times 年龄_{it}+\varepsilon_{it} \quad [4.12]$$

尽管方程4.12描述的模型不能被估计，但它可以用年龄—时期—世代效应来证明同号原则的含义。把方程4.12中的年龄用（时期—世代）代替，得到：

$$Y_{it}=\beta_0+(\beta_P+\beta_A)\times 时期_{it}+(\beta_C-\beta_A)\times 世代_{it}+\varepsilon_{it} \quad [4.13]$$

比较方程4.13和方程4.4，我们得到以下重要结果：

$$世代内的斜率：\beta_1=\beta_P+\beta_A \quad [4.14]$$

$$世代间的斜率：\beta_2=\beta_C-\beta_A \quad [4.15]$$

在考虑年龄和世代效应的符号时，记住年龄和世代是从相反的方向来编码的。在任一给定的时点，那些在年龄上编码较低的人，在世代上编码则较高，因为他们出生得更晚。

从方程4.14和方程4.15中，我们可以很直接地判断出什么时候 β_1 和 β_2 会在线性—叠加模型中有相同的符号。如果 $\beta_C>\beta_A$，世代间的斜率 β_2 则为正；如果 $\beta_C<\beta_A$，世代间的斜率 β_2 则为负。因此，在个人变化为正的情况下，即 $\beta_1>0$ 时，那么在且只有在 $\beta_C>\beta_A$ 的情况下，总变化才快于个人变化。这个条件没有对年龄和世代效应的符号进行任何限制，它们可以同时为负，只要在数轴上，β_A 在 β_C 的左边，世代间的斜率就为正（$\beta_A<\beta_C$）。在个人变化为负的情况下，即 $\beta_1<0$ 时，那么在且只有在 $\beta_C<\beta_A$ 的情况下，总变化才快于个人变化。同样，它对 β_C 和 β_A 的符号没有任何限制，重要的只是它们在数轴上的相对位置。

总而言之，在线性—叠加的标准假定下，当总变化和个人变化同向时，在且只有在以下情况下，总变化会快于个人变化：如果是上升趋势，则为 $\beta_C>\beta_A$；如果是下降趋势，则为 $\beta_C<\beta_A$。

斜率符号相同原则的实例：关于性别角色的态度

近几十年来，大量女性参加有薪工作，这是一种很可能会对年轻人的态度产生巨大影响的社会变迁。与莱德的研究相对应，在美国人对性别角色的态度

的问题上，人们预期世代效应和时期作用将会互补。

为了检验这些预期，我把美国综合社会调查中重复得最多的 4 个关于性别角色的问题作为分析的数据（表 4.2）。第一个问题是关于女性的有薪工作（WORK），接下来的两个问题是关于女性的政治参与（PRES，POLI），第四个问题似乎同时包含政治和“家庭生活”的维度（HOME）。表 4.2 中的“加权”指为美国综合社会调查中少计的已婚人士而进行的调整。[7]然而加权没有产生任何影响，所以在后面的表格中，我只汇报没有加权的结果。

1972 年，65%的受访者说，即使女人的丈夫可以支持她的生活，他们仍然认可她去参加有薪工作；1988 年，80%的人表示支持此行为（表 4.2）。相似的变化模式也出现在其他关于女性角色的问题上（有趣的是，这些百分比并不因受访者的性别而异）。1972 年，74%的人说他们会投票支持一位合格的女性提名人当总统；1988 年，88%的人说他们会这么做。1974 年，略高于半数的受访者不同意“大多数男性在性情上比大多数女性更适合参与政治”；1988 年，大约 2/3 的人不同意这个说法。那些不同意“女性应该照顾家庭，把治理国家让给男人”的人从 1974 年的 64%上升到 1988 年的 79%。

表 4.2　美国对性别角色的态度转变

	1972 年均值[a]	1988 年均值	变化
WORK：如果已婚女性的丈夫可以支持她的生活，你同意还是不同意她从生意或工业工作中赚钱？（1 = 同意，0 = 不同意）			
不加权的样本	0.654	0.804	0.150*
加权的样本	0.652	0.807	0.155*
PRES：如果你的政党提名一位女性为总统，且她能做好该工作，你是否会投票给她？（1 = 会）			
不加权的样本	0.737	0.879	0.142*
加权的样本	0.737	0.879	0.142*
POLI：告诉我你同不同意这句话：大多数男性在性情上比大多数女性更适合参与政治。（1 = 不同意）			
不加权的样本	0.532	0.667	0.135*
加权的样本	0.530	0.670	0.140*

续表

	1972 年均值[a]	1988 年均值	变化
HOME:你同不同意这句话:女性应该照顾家庭,把治理国家让给男人。(1 = 不同意)			
不加权的样本	0.644	0.788	0.144*
加权的样本	0.643	0.791	0.148*

注:a. POLI 和 HOME 为 1974 年的均值。
*表示 $p < 0.0001$。

总变化的速度是否快于一般成年人态度的改变速度呢?表 4.3 为 4 个关于性别角色的问题汇报了世代内和世代间的斜率的估计。对这些问题的回答是二分变量,所以在统计上我们需要用 logistic 回归。对于这 4 个问题,世代内和世代间的斜率都同号(普通最小二乘回归得出的结果是一样的)。根据刚刚推导出来的条件,我们得知,关于性别角色的观念在成人世界中的整体变化速度要比一个典型的成年人变化得快。

表 4.3 世代内和世代间的对性别角色的态度改变的估计:logit 系数

性别角色测量	世代内($\hat{\beta}_1$)	世代间($\hat{\beta}_2$)
WORK 1972—1988(N = 14376)	0.029*	0.032*
PRES 1972—1988(N = 14188)	0.029*	0.031*
POLI 1974—1988(N = 11058)	0.029*	0.024*
HOME 1974—1988(N = 12665)	0.029*	0.035*

注:* $p < 0.01$。

表 4.4 活着的成年人对性别角色的态度改变的估计(1972[1974]—1988 年)

	变化	
性别角色问题	总变化[a]	活着的人的变化[b]
WORK	0.150	0.087
PRES	0.142	0.089
POLI	0.135	0.036
HOME	0.144	0.065

注:a. 从表 4.2 而来。
b. 活着的成年人的平均变化。计算方法参见正文。

代数分解法得出的结论相同。对这 4 个问题中的任何一个，支持性别平等的回答在 1988 年比在 1972 年或 1974 年多 15%。[8]我们可以通过代数分解来将这 15 个百分点的变化与活着的成年人的变化做比较。

因为想知道活着的人的变化，我们用世代在时点 2 时所占总体的比重对世代内的变化 $\mu_{j2}-\mu_{j1}$ 进行加权，而不是用世代在时点 1 时所占总体的比重。我们用 $\sum_j p_{j2}(\mu_{j2}-\mu_{j1})$ 来估计活着的一般成年人的变化。[9]

表 4.4 汇报了结果。对所有 4 个问题，活着的人的平均变化明显慢于总变化(因为活着的人的均值距离最大值比所有成年人的均值距离最大值远，活着的人更慢的变化不能归因于天花板效应)。总的来说，那些支持"在已婚女性的丈夫可以支持她的生活的情况下，女人仍可以从生意或工业工作中赚钱"的人增加了 15 个百分点，而对活着的成年人来说，则增加了 8.7 个百分点。1988 年，88%的人说，如果他们的政党提名一个合格的女性做总统，他们会投票支持，这比 1972 年增加了 14%；对仍然生存的成年人来说，这个百分比则增加 8.9%。其他问题表现出相似的变动。

简言之，表 4.4 的结果与斜率同号原则的预测一致。在美国，对性别角色的态度在成年人社会中的变化总的来说比典型的成年人快。

总结

如何把微观过程和宏观过程联系起来是社会学一直在讨论的一个话题(Alexander, Geisen, Munch & Smelser, 1987; Coleman, 1986; Durkheim, 1895/1938)。线性—叠加例子中的斜率同号原则与这种联系有关，它给为什么总变化有时候快于个人变化提供了线索(当然，在现实世界中，作用可能是非线性的和交互的[Elder, 1974]，然而，线性—叠加的理想类型是一个很有用的起点)。这个线索就是，当世代内和世代间的斜率有相同的符号时，成人世界的变化快于其中一般成年人的变化。用年龄和世代效应来表述，斜率符号相同的原则意味着在上升趋势下，$\beta_C > \beta_A$，在下降趋势下，$\beta_C < \beta_A$。在这两种情况下，社会的变化都快于个人的变化。

第 5 章 | 分解总变化的一般模型

上一章介绍了两种从社会变迁中分离世代更替作用的方法。在隔离世代更替成分的时候，我们关注的只是我们所观察到的变化发生的来源，这些变化来源于人口更替还是个人变化？由于目的是确定变化的来源，所以这个分解方法并不通过加入变量来解释为什么个人发生了变化，或者为什么不同的世代有所不同。

本章用因果变量来回答变化的问题。目的是展示一个一般模型，该模型用其他变量自身和作用在时点 1 到时点 2 的变化，来解释一个变量在相同时间内的总变化。这个模型假定只有两个测量时点(而线性分解中有 T 个时点)。因为这个模型在形式上与人们熟知的回归标准化模型一样，所以在此没有必要用新的分析来展示它，而只需要依赖已有的研究。我使用的研究案例是在政治科学文献中 1960 年后美国选举投票比例下降的奇怪现象。

模型

我们来考虑简单的线性回归模型，Y 是 X 的一个函数：

$$Y = \alpha + \beta X + \varepsilon \qquad [5.1]$$

如前所述(方程 4.5)，在传统的假定 $E(\varepsilon) = 0$ 下，Y 的均值是它的预期值：

$$\begin{aligned} \overline{Y} &= E(Y) \\ &= E(\alpha + \beta X + \varepsilon) \\ &= \alpha + \beta \overline{X} \end{aligned} \qquad [5.2]$$

方程 5.2 指出，回归线会穿过点($\bar{X}$, $\bar{Y}$)。因此，我们可以用 X 的均值来得到 Y 的均值，即用 X 的均值乘以斜率再加上截距。此处更重要的是，方程 5.2 告诉我们，Y 的均值可以用以下 3 个部分表示出来：(1)X 的均值；(2)$X-Y$ 关系的斜率；(3)截距。

因为 $\bar{Y}$ 可以用 $\bar{X}$、回归斜率和截距来表示，所以 $\bar{Y}$ 的变化可以用 $\bar{X}$ 的变化、斜率的变化和截距的变化来表示。加入下脚标来表示时间，由方程 5.2 可得 Y 的均值的变化为：

$$
\begin{aligned}
\bar{Y}_2-\bar{Y}_1 &= (\alpha_2+\beta_2\bar{X}_2)-(\alpha_1+\beta_1\bar{X}_1) \\
&= (\alpha_2-\alpha_1)+(\beta_2\bar{X}_2-\beta_1\bar{X}_1) \\
&= (\alpha_2-\alpha_1)+(\beta_2\bar{X}_2-\beta_1\bar{X}_1)+(\beta_2\bar{X}_1-\beta_2\bar{X}_1) \\
&\quad +(\beta_1\bar{X}_2-\beta_1\bar{X}_2)+(\beta_1\bar{X}_1-\beta_1\bar{X}_1) \\
&= (\alpha_2-\alpha_1)+(\beta_2-\beta_1)\bar{X}_1+\beta_1(\bar{X}_2-\bar{X}_1)+(\beta_2-\beta_1)(\bar{X}_2-\bar{X}_1) \\
&= \Delta\alpha+\Delta\beta\bar{X}_1+\beta_1\Delta\bar{X}+\Delta\beta\Delta\bar{X}
\end{aligned}
\tag{5.3}
$$

总而言之，$\bar{Y}$ 的变化可以分解为 4 个部分，分别反映在截距的变化($\Delta\alpha$)、解释变量的均值($\Delta\bar{X}$)的变化和解释变量的作用($\Delta\beta$)的变化上(其他分解方法也是可能的，把方程 5.3 中的 4 个部分用不同的方式相加，可以变成 3 个部分或者 2 个部分)。我们可以发现，与第 4 章所讲的回归方法不同的是，这里的分解是完全的，这些部分加起来等于 $\bar{Y}$ 的变化。

熟悉回归标准化的读者会发现，方程 5.3 与回归标准化的分解方程的形式是一样的(Sobel, 1983：公式 4)。不同之处在于，这里的下脚标 1 和 2 表示时间，而不是群体。回归标准化，正如它的名字所暗示的那样，是用回归来标准化分布的方法。通常来说，其目的是将两个群体的分布标准化，以去除由于构成不同而造成的两个群体在因变量上的差异。举例来说，研究收入上的性别歧视经常运用回归标准化法来调整男女工人在年龄、每周工作时间、工作经验等方面的不同。

因为我所提出的分解总变化随时间变化的模型(方程 5.3)在形式上与用于

分解两个群体之间均值差异的回归标准化法一样，回归标准化法的内在问题也存在于用方程5.3来分解变化上。这点很重要，因为回归标准化法受到“起点依赖”或“位置变化”问题的困扰（Clogg & Eliason, 1986; Firebaugh, 1992; Jones & Kelley, 1984）。问题是这样的：给 X 加入一个常数，我们就改变了方程5.3前两部分的大小。实际上，除非 X 有可靠的零点，否则只要改变 X 的起点，即给 X 加入一个常数，截距的改变（$\Delta\alpha$）对总变化的贡献就可以变成任意大小。

为了更好地理解这个问题，请首先注意方程5.3的第三部分和第四部分没有起点依赖的问题。第三部分 $\beta_1\Delta\overline{X}$ 是 Y 的均值的总变化中完全来自解释变量的均值变化的部分。第四部分 $\Delta\beta\Delta\overline{X}$ 是总变化中来自斜率变化和均值变化的共同作用的部分。这些部分没有起点依赖的问题，因为给 X 加一个常数不会影响 X 均值的差 $\Delta\overline{X}$。然而，给 X 加一个常数会改变 X 的均值，所以除非 $\Delta\beta=0$，否则就等于给第二部分 $\Delta\beta\overline{X}_1$ 增加了一个恒定的作用。由此可知，增加一个常数也会影响 $\Delta\alpha$，因为以下方程必须平衡：由方程5.3，$\Delta\alpha=\Delta\overline{Y}-(\Delta\beta\overline{X}_1+\beta_1\Delta\overline{X}+\Delta\beta\Delta\overline{X})$ 可知，给 X 加一个常数，不会影响 $\Delta\overline{Y}$、$\beta_1\Delta\overline{X}$ 和 $\Delta\beta\Delta\overline{X}$（但是 $\Delta\beta\overline{X}_1$ 会受到影响），所以很显然，$\Delta\alpha$ 也会受到影响。

简言之，除非 X 的作用是恒定的（$\Delta\beta=0$），否则 $\Delta\alpha$——Y 的变化中不能被 X 自身及作用的变化所解释的部分——可以通过给 X 加入一个常数使之向对研究者有利的方向改变。例如，通过给 X 加入一个适当的常数，我们可以使 $\Delta\alpha=0$，然后宣称 $\overline{Y}$ 的变化全部来自 X 自身及作用的变化。或者我们可以给 X 加一个常数，使截距的差别等于 $\overline{Y}$ 的差别（即我们可以使其他几部分加起来为0）。因此，用一样的数据和变量，一位研究者可以声称，X 自身及作用的变化可以解释所有 Y 的均值的变化（$\Delta\alpha=0$），而另一位研究者可以声称，X 自身及作用的变化完全不能解释 Y 的均值的变化（$\Delta\alpha=\Delta\overline{Y}$，因此与 X 有关的部分的贡献相加后为0）。

有两种方法可以解决这个问题。最好的方法就是选择有非任意零点的自变量（X），也就是选择定比层次的测量。在定比变量里，零点是固定的，截距的位置也固定，所以可以有唯一的分解方法。如果不能得到定比的测量，研究者可以合并方程5.3的前两部分，得到以下的几个部分（Sobel, 1983）：$\Delta\alpha+$

$\Delta\beta\overline{X}_1$，参数变化的贡献(注意，这里用了不同的术语"参数变化"，指的是截距的变化以及斜率的变化)；$\beta_1\Delta\overline{X}$，$X$ 自身的变化的贡献；$\Delta\beta\Delta\overline{X}$，$X$ 自身及作用的变化的贡献。

这里的底线是，区分截距改变的作用和斜率改变的作用要求有定比的测量。至少我们要注意到，当截距可以任意移动时，短语"截距变化的作用"没有实质含义，这是该方法的一个缺陷。当截距的位置是任意的时候(正如在以定距测量自变量 X 时)，$\Delta\alpha$ 不过是用来使方程 5.3 平衡的项。

多变量分解

将方程 5.3 的分解扩展到多个解释变量的情况也很直接。为了便于说明，用向量来标记更简便。用 $\mathbf{X}$ 表示一个 $1\times q$ 的自变量的行向量，用 β 表示一个 $q\times 1$ 的参数的列向量。由此 Y 的均值可以表示为 $\overline{Y}=E(\alpha+\mathbf{X}\beta+\varepsilon)=\alpha+\overline{\mathbf{X}}\beta$，此处 $\overline{X}$ 是一个 $1\times q$ 的 X 的均值的向量。再次用下脚标来表示时间，可得 Y 的均值从时点 1 到时点 2 的变化为：

$$\begin{aligned}\overline{Y}_2-\overline{Y}_1&=(\alpha_2+\overline{\mathbf{X}}_2\beta_2)-(\alpha_1+\overline{\mathbf{X}}_1\beta_1)\\&=(\alpha_2-\alpha_1)+\overline{\mathbf{X}}_1(\beta_2-\beta_1)+(\overline{\mathbf{X}}_2-\overline{\mathbf{X}}_1)\beta_1+(\overline{\mathbf{X}}_2-\overline{\mathbf{X}}_1)(\beta_2-\beta_1)\\&=\Delta\alpha+\overline{\mathbf{X}}_1\Delta\beta+\Delta\overline{\mathbf{X}}\beta_1+\Delta\overline{\mathbf{X}}\Delta\beta\end{aligned}\qquad[5.4]$$

我们不应该把方程 5.4 和第 4 章介绍的线性分解法混淆。首先，方程 5.4 的模型是基于两个截面数据，而第 4 章的模型使用所有的截面数据。第二，方程 5.4 的分解模型允许解释变量的作用发生变化，而第 4 章的模型假定参数(包括截距)不随时间发生变化。如果我们假定方程 5.4 中 $\Delta\alpha=\Delta\beta=0$，则模型可以简化为第 4 章模型中的 $\Delta\overline{Y}=\Delta\mathbf{X}\beta$ 形式(参见方程 4.7)。第三，方程 5.4 的分解法是一个灵活的模型，可以有不固定数量的自变量。相反，第 4 章的分解法是根据特定的目的而设计，即从世代内变化的作用中区分世代更替的作用，所以相同的自变量——测量年份和出生年份——总是被使用。

例子：美国选举中投票比例的下降

为了演示方程 5.4 所提供的分解框架的可能应用，我们来分析在过去几十年中，美国的全国选举中投票比例的变化趋势。美国的投票率在 1960 年肯尼迪—尼克松总统选举过后的 30 年间都在下降，除了在 1992 年的选举中出现过反弹(Abramson et al.，1994)。从个人层次影响投票率的因素的变化来看，这个长时期的投票率下降特别有趣。举例来说，选民的教育水平自 1960 年后都在上升，并且我们知道，正式教育与投票可能性之间有很强的正向关系(Wolfinger & Rosenstone，1980)。根据教育的变化趋势，我们预测投票率会上升，而不是下降。

有极为丰富的实证研究文献(回顾请参见 Abramson et al.，1994：第 4 章)试图解答这个美国投票率的"谜题"(Brody，1978)，即在选民越来越多、受教育程度越来越高且登记为选民的要求越来越低的情况下，投票率却越来越低的谜题。美国对投票参与进行了长时间的且令人敬佩的实证研究(例如，Campbell，Converse，Miller & Stokes，1960；Kleppner，1982；Merriam & Gosnell，1924；Verba & Nie，1972；Wolfinger & Rosenstone，1980)，这些对投票率谜题的研究受到了这一传统的启发。对投票参与的研究得出了一系列与投票行为有关的人口学因素和态度因素(Bennett & Bennett，1987)，从这一系列因素中，艾布拉姆森和奥尔德里奇认为，其中两个因素与投票率下降最有关系："投票参与的下降主要来自两个基本的态度变化趋势：美国选民对政党忠诚的弱化和对政府反应的不信任"(Abramson & Aldrich，1982：502)。特谢拉的全面分析得出了相同的结论(Teixeira，1987、1992)，并指出另一点与投票率下降联系在一起的因素，即"社会联系的明显下降表现为选民更年轻、更少人结婚、更不常参加礼拜"(Teixeira，1992：57)。

艾布拉姆森和奥尔德里奇的研究与特谢拉的研究得出的结论是一致的，这点值得我们注意，因为他们使用了不同的分解方法。他们都是用累积的全国选举研究数据(开始于 1960 年)，并用虚拟变量来表示选举年，研究投票参与随时

间的变化。然而，艾布拉姆森—奥尔德里奇分析使用斜率变化模型以允许两个解释变量的作用可以随时间而变化。特谢拉使用了更多的解释变量，但是(除了一些补充的回归以外)假定它们的作用是随时间不变的，也就是说，他假定 $\Delta\beta = 0$。把 $\Delta\beta = 0$ 代入上面的方程 5.4，我们可以更好地理解特谢拉分解任何两次选举中投票率不同的方法：

$$\begin{aligned}\overline{Y}_2 - \overline{Y}_1 &= \Delta\alpha + \overline{\mathbf{X}}_1\Delta\beta + \Delta\overline{\mathbf{X}}\beta_1 + \Delta\overline{\mathbf{X}}\Delta\beta \\ &= \Delta\alpha + \Delta\overline{\mathbf{X}}\beta_1\end{aligned} \qquad [5.5]$$

换言之，特谢拉的方法把投票率下降分解为 $\Delta\overline{\mathbf{X}}\beta_1$，即构成效应(来自自变量 X 的均值的变化部分)，以及 $\Delta\alpha$，即不能被自变量 X 解释的变化。

在分解投票率下降时，特谢拉检验了把自变量 X 加入模型后截距($\Delta\alpha$)的变化(模型中 $\Delta\alpha = (\overline{Y}_2 - \overline{Y}_1) - \Delta\overline{\mathbf{X}}\beta_1$)(Teixeira，1987、1992)。“如果加入一个变量后，这些系数的大小显著地下降，这就意味着那个变量的分布变化对投票率下降有重要影响。”(Teixeira，1987:45；1992:196)将方程 5.5 和方程 5.4 相比较，我们注意到，这个解释假定 $\mathbf{X}$ 的作用是恒定的($\Delta\beta = 0$)。如果 $\mathbf{X}$ 的作用随时间不变，我们可以解释说，从时点 1 到时点 2 的截距的变化是 Y 的均值的变化中不能被解释变量说明的部分。[10]

作为一种与特谢拉分析投票率下降的叠加分解模型不同的模型，我们可以用方程 5.4 的模型。方程 5.4 的模型更为一般化，因为它包含了解释变量作用改变的贡献(如果有)。考虑从 1960 年到 1988 年 8 次总统选举投票率的下降。我们可以把这个下降用一步分解为：

$$\overline{Y}_{88} - \overline{Y}_{60} = \Delta\alpha + \overline{\mathbf{X}}_{60}\Delta\beta + \Delta\overline{\mathbf{X}}\beta_{60} + \Delta\overline{\mathbf{X}}\Delta\beta_{60} \qquad [5.6]$$

换言之，我们只用 1960 年和 1988 年的数据进行分解，忽略期间 6 次选举的数据(重复一下，如果一些自变量 X 不是定比层次的，方程 5.6 中的第一项和第二项就需要合并)。或者我们可以相继地分解变化以使用全部 8 次选举数据——首先分解从 1960 年到 1964 年的变化，然后是 1964 年到 1968 年的变化，如此类推——之后再相加。我推荐第二种方法，尽管它更费力。

总结

本章展示了一个一般的模型，它可以把 Y 的总变化分解为解释变量自身的变化（构成作用）、解释变量的作用的变化和截距的变化。这个模型可以作为研究社会变迁的样板。然而，研究者不应该机械地使用这个方法。分解方法基于它所使用的解释变量，只有在解释变量有意义的时候，分解才是有意义的。举例来说，我们可以把分解方程应用到任何变量上，只要这些变量跟投票有关，并且在过去 30 多年中表现出向上或者向下的趋势，以“解释”投票率下降的问题。这样机械的运用方法也许能得到统计上显著的结果，但却不能告诉我们社会世界的任何有用信息。

第 6 章 | 发掘个人层次关系的变化

本书的内容是分析社会变化。迄今为止，我关注的是总变化，然而术语“社会变迁”有时候指的是个人层次关系的变化。本章将介绍如何对重复调查数据使用简单的回归模型来检验解释变量的作用是否随时间改变。

参数变化模型

在本章中，术语“参数变化模型”指的是用于研究 $\Delta\beta$，即自变量 X 的作用变化的模型。正如上一章介绍的那样，作为总变化的一个部分，研究者也可以研究参数的变化。然而，在本章中，我关注对参数变化本身的研究，而不是研究他们对更大的社会变迁的贡献。因为个人层次作用的变化本身就很有趣，所以可以独立使用参数变化模型来分析。

我用交互项来反映 X 作用的变化。如果一个变量对另一个变量的作用受到第三个变量 Z 的层次的影响，就出现了交互作用。我们说 X 的作用随时间变化，也就是说，X 的作用受时间影响。这是从经典意义上来理解交互作用这个术语的例子。

我所用的交互项的形式是 X 乘以 D_{YR}，其中 D_{YR} 是一个虚拟变量，当它指某一年(或者月、周，取决于测量的频率)时，取值为 1。正如我们将看到的，解释变量 X 既可以是连续的，也可以是分类的。

假设我们想知道，在美国，决定人们政党身份的因素在 1994 年与 20 多年前是不是一样，我们可以做两个平行的分析——一个使用 1974 年的数据，一个使用 1994 年的数据——来比较系数。用这种方法立刻可以很明显地看到哪些差

别是显著的，哪些是不显著的。为了检验来自个别样本(这里是个别年份)的样本系数的不同是否在统计上显著，我们必须进行合适的显著性检验。一种方法是对各个样本分别进行回归分析，再用朗和麦尔德给出的方程对我们感兴趣的差别进行 t 检验(Long & Miethe, 1988:125—129)。然而在有重复调查数据的情况下，这个方法不必要，因为数据搜集者经常把新的样本和以前的样本合并。有了这样的累积数据，通过增加一个交互项来估计一个模型更容易检验参数的变化，这就是我在这里要介绍的方法。

模型的一般形式

用向量来表示，参数变化模型的形式为：

$$E(Y) = \alpha + \gamma D_{YR} + \mathbf{X}\beta + (\mathbf{X}D_{YR})\delta \qquad [6.1]$$

D_{YR}是表示年份的虚拟变量，$\mathbf{X}$ 是 D_{YR}以外的一组预测变量的向量，β 和 δ 是参数的向量。如果在向量 $\mathbf{X}$ 中有 q 个变量，那么 $\mathbf{X}$ 和 $\mathbf{X}D_{YR}$各有 $1\times q$ 的维度，且 β 和 δ 各有 $q\times 1$ 的维度。

方程 6.1 用最一般的形式来表示模型，每个自变量都有对应的交互项。我只是为了展示出这个模型的适用范围。在使用这个模型时，研究者经常会只给一个子集中的自变量 X 加上交互项，尤其是当自变量很多的时候。给模型加交互项的原则与给模型加自变量的原则是一样的，即必须有好的理由(理论、已有的证据或者有理的论据)以使人相信这个变量是有作用的。滥用参数变化模型的最大危险在于，研究者“漫无目的地探寻”随时间变化的作用。这类交互项只有在我们相信自变量的影响随时间而变化时才加入模型。如果交互项被随意地与很多自变量一起加入模型，样本系数的有些变化可能只是因为偶然的原因而在统计上显著，参数的真正变化却可能由于多重共线性导致标准误变大而变得不显著。

从具体的事例来考虑，通常都可以帮助我们思考。请考虑从 1974 年到 1994 年决定人们政党身份的因素有没有发生改变这个例子。美国综合社会调

查包含以下的问题："总的来说，你一般把自己看做共和党人、民主党人、独立人士，还是其他？"编码从 0 = 很强的民主党人（认同感）到 6 = 很强的共和党人（认同感）。如果假定在这个问题中，独立人士（编码为 3）作为一个单独的维度处于中间位置，我们就可以把它作为测量共和党人身份认同强度的变量。为了与后面的介绍一致，如果我们改变编码方向为 6 = 很强的民主党人（认同感），并用 1974 年和 1994 年的美国综合社会调查数据来估计这个模型，那么：

$$E(Dem) = \alpha + \gamma D_{94} + \mathbf{X}\beta + (\mathbf{X}'D_{94})\delta \qquad [6.2]$$

此处，D_{94}是一个虚拟变量，当年份是 1994 年时，取值为 1，$\mathbf{X}$ 是一个 $1 \times \mathrm{q}$ 的变量的向量，被认为会影响对民主党的政党认同，$\mathbf{X}'$是一个 $1 \times \mathrm{p}$ 的变量的向量，它包含了 $\mathbf{X}$ 中的一部分变量（或者当 $\mathrm{p} = \mathrm{q}$ 时，$\mathbf{X}' = \mathbf{X}$）。参数 α 是 1974 年时的 y 的截距，$\alpha + \gamma$ 是 1994 年的截距。向量 β 包含 q 个预测变量 $X_1, X_2, \cdots, X_p, \cdots, X_q$，对应的参数是 $\beta_1, \beta_2, \cdots, \beta_p, \cdots, \beta_q$。$X_{p+1}, X_{p+2}, \cdots, X_q$ 对应的 β 表示预测变量的直接作用，它们被认为在 1994 年时的作用与在 1974 年时相同。与之不同的是，$X_1, X_2, \cdots, X_p$ 对应的 β 表示预测变量在 1974 年的直接作用，这些变量的作用被假设为从 1974 年到 1994 年发生了改变。所以，β_1 是 1974 年 X_1 对民主党身份认同的直接作用，β_2 是 1974 年 X_2 对民主党身份认同的直接作用，等等，而 $X_1, X_2, \cdots, X_p$ 被假设为在 1974 年的作用与在 1994 年的作用不同。

请注意，参数变化模型允许 X 的作用随时间变化，但是不包含其他类型的交互项。例如，在运用方程 6.2 的时候，我们可能检验地区（南方对非南方）和教育对民主党身份认同的作用，看地区和教育对政党身份认同的作用在近几十年是否发生了变化。在使用方程 6.2 的时候，我们假定教育对政党身份认同的作用在南方和非南方是一样的。当然，这个假设可以通过加入地区 × 教育这个交互项来检验，此处地区是一个虚拟变量。因为在参数变化模型中加入这样一个交互项很直接，所以我认为没必要在这里加入这一项把标记弄复杂（不熟悉这种交互项的读者可以参看标准的统计教材，例如 Agresti & Finlay, 1986）。参数变化模型的要点在于检验 X 的作用是否随时间变化，这是我们此处的关注点。

作用变化的显著性检验

在这里,方程 6.2 的交互项($\mathbf{X}'D_{YR}$)δ 是重点,因为它显示了 X 的作用是否随时间变化。向量 $\mathbf{X}'D_{94}$ 包含变量 X_1D_{94}, X_2D_{94}, …, X_pD_{94},此处 X_1D_{94} 是 1994 年对应的 X_1 且在 1974 年为 0, X_2D_{94} 对应的是 1994 年的 X_2 且在 1974 年为 0,等等。因此,向量 δ 表示的是 1974 年到 1994 年间 X_1, X_2, …, X_p 的作用的变化(例如,δ_1 是 X_1 作用的变化)。因为从 β_1 到 β_p 反映的是自变量 X 在 1974 年的直接作用,由此可得 X'_1 在 1994 年的直接作用是 $\beta_1+\delta_1$, X'_2 在 1994 年的直接作用是 $\beta_2+\delta_2$,等等。

我们可以检验单一交互项的统计显著度,也可以同时检验一组交互项。单独系数的检验用 t 检验,同时检验多个交互项用 F 检验(Long & Miethe, 1998)。SPSS 和 SAS 之类的程序软件包通常都会给出各个系数的 t 值。要同时检验多个交互项,我们须再次估计 X 的作用,但这次不包含交互项:

$$E(Dem)=\alpha_r+\gamma_r D_{94}+\mathbf{X}\beta_r \qquad [6.3]$$

此处,下脚标 r 用来区分简化模型中的参数和完整模型中的系数,简化模型如方程 6.3 所示,完整模型如方程 6.2 所示。恰当的 F 检验为:

$$F=[(\mathrm{SSE}_r-\mathrm{SSE}_c)/\mathrm{p}]/[\mathrm{SSE}_c/(N-\mathrm{k})] \qquad [6.4]$$

此处,SSE_r 是在简化模型中的误差平方和,SSE_c 是在完整模型中的误差平方和,k 是完整模型中的参数的数量,p 是交互项的数量,N 是样本量(方程 6.4 可以重写为用 R^2 的差来表示的形式,有些读者可能对这种形式更熟悉。这也是有些统计软件,譬如 SPSS 使用的形式)。如果一组交互项不能减少误差平方和,那么 $\mathrm{SSE}_c=\mathrm{SSE}_r$,且 $F=0$,所以 F 在这种情况下是检验零假设,即交互作用共同为 0,也就是 $\delta_1=\delta_2=\cdots=\delta_p=0$(Agresti & Finlay, 1986:456)。如果我们不能拒绝这个零假设,我们就认为自变量 X 在 1974 年的作用与在 1994 年的作用一样(在 logistic 回归的情况下,显著性检验是基于卡方检验而不是 F 检验)。

两个简单的例子

用交互方法来研究个人层次关系的变化很灵活。自变量 X 可以是连续的，也可以是分类的，也可以两者都有。对连续的 X，δ 是斜率的变化。对分类的 X，δ 是分类均值的变化（或者在其他变量是连续变量的情况下，δ 是截距的变化）。

在 X 为分类变量的情况下，我们假设从 1974 年到 1994 年民主党身份认同的地区差异（南方对非南方）变小了。我们可以通过以下合并了 1974 年和 1994 年数据的模型来检验假设：

$$E(Dem) = \alpha + \gamma D_{94} + \beta \times \text{地区} + \delta \times \text{地区} \times D_{94} \qquad [6.5]$$

此处，地区是一个虚拟变量，当（受访者）居住在南方时取值为 1。因为 D_{94} 在 1974 年时取值为 0，所以 β 是 1974 年的区域差异，δ 是 1994 年和 1974 年的区域差异之差。因此，$\delta = 0$ 意味着 1994 年和 1974 年的区域差异一样。如果 δ 和 β 不为 0，当 δ 和 β 同号时，1994 年的区域差异（比 1974 年）更大，当 δ 和 β 异号时，1994 年的区域差异更小（除非在 δ 在绝对值上是 β 的两倍以上）。无论地区的编码是 1 = 南方或者 1 = 非南方，这个结论都成立。

现在我们来考虑 X 为连续变量的情况。如果我们假设，在美国教育和民主党身份认同的双变量关系发生了改变，我们可以通过以下模型来检验这个假设：

$$E(Dem) = \alpha + \gamma D_{94} + \beta \times \text{教育} + \delta \times \text{教育} \times D_{94} \qquad [6.6]$$

正如在方程 6.5 中的情况那样，相同的原则在方程 6.6 中同样成立。当 δ 和 β 同号时，教育在 1994 年的作用比在 1974 年更大；当 δ 为 0 时，教育的作用不变（1974 年的斜率与 1994 年一样）；当 δ 和 β 异号时，教育在 1994 年的作用比在 1974 年更小，除非 δ 的大小是 β 的两倍以上。方程 6.6 和方程 6.5 之间的重要差别是，方程 6.6 中的参数表示的是斜率和斜率之差。在方程 6.6 中，β 是 1974 年的教育的斜率，而 δ 是 1994 年与 1974 年的教育的斜率之差。

参数变化分析的步骤：种族与民主党身份认同

文献中经常提及美国黑人比白人更可能参与民主党并投票给民主党候选人(Abramson et al.，1994:第5章)。也有证据表明,参与民主党的种族差异正在扩大(Abramson et al.，1994:第8章)。在这里,我使用1994年的美国综合社会调查数据和在那之前20年的美国综合社会调查数据,运用参数变化模型来检验在美国参与民主党的种族差异在扩大这个假设。

我从只有少数变量的模型开始。在这里,越简单越好。我想说明的重点是方法性的,我不想读者由于大量变量而忽略了这些方法上的重点。简单的模型就足以展示用参数变化模型分析重复调查的要点。一旦理解了这些要点,正如我所展示的那样,增加协变量以改善模型就是很容易的事。读者要记住,这个例子的主要目的是解释方法,而不是发展一个实质领域。

民主党身份认同的种族差异

表6.1汇报了在1973/1974年和1994年的美国综合调查中,白人和黑人受访者各自认为自己有"很强的民主党人(认同感)"的百分比。这些百分比是来自有政党身份、种族和其他4个变量等数据的受访者($N=5581$),这4个变量分别是教育、地区、性别和年龄,它们将在以后作为协变量被纳入回归分析。因为1994年的美国综合社会调查的样本量为通常的两倍(在1994年美国综合社会调查改为两年调查一次),我合并了1973年和1974年的调查。为了方便起见,我用1974年的调查指代1973年和1974年的调查。

表6.1 民主党身份认同的种族差异[a](1974年对1994年)

种族	百分比		比数[b]	
	1974	1994	1974	1994
白人	14.71	10.48	0.1725	0.1171
黑人	33.53	40.97	0.5044	0.6941

续表

种　　族	百分比		比　数[b]	
	1974	1994	1974	1994
百分比差别(白人—黑人)	−18.82	−30.49		
比数比(白人/黑人)	0.342	0.169		
1974 年至 1994 年种族差异的变化				
以百分比表示：−30.49−(−18.82) = −11.67%				
以比数表示：0.169/0.342 = 0.494				

注：a. 自我认定为“很强的民主党人(认同感)”。样本量为 5581(1973/1974 年为 2806，1994 年为 2775)。样本包含所有在种族、政党身份和 4 个之后加入的协变量上没有缺失值的受访者，这 4 个协变量是教育、地区、性别和年龄。

b. “比数”定义为 $p/(1-p)$，这里 p 是概率。因为 p 等于百分数/100，比数在这里是由百分比计算而来：比数 = 百分数 /(100 − 百分数)。

为了方便进行 logistic 回归结果的比较，表 6.1 也报告了民主党身份认同的比数。表 6.1 报告的比数由百分数计算而来：比数 = 百分数 /(100 − 百分数)。例如，在 1974 年，14.71%的白人受访者认为自己有很强的民主党人认同，其比数为 0.1725(14.71[100 − 14.71] = 0.1725)。

表 6.1 的结果显示，从 1974 年到 1994 年，民主党身份认同的种族差异变大了。1974 年，白人相对于黑人民主党人身份认同的“赤字”为 18.82%：14.71%的白人认为自己有很强的民主党人认同，相对的，有 33.53%的黑人这么认为。1994 年，这一赤字激增到 30.49%：10.48%的白人认为自己有很强的民主党人认同，但有 40.97%的黑人这么认为。

从比数的角度说，白人认为自己有很强的民主党人认同的比数从 1974 年的 0.1725 下降到 1994 年的 0.1171，而非洲裔美国人的比数从 1974 年的 0.5044 提高到 1994 年的 0.6941。白人/黑人比数比——白人认为自己有很强的民主党人认同感的比数与黑人认为自己有很强的民主党人认同感的比数之间的比率——从 1974 年的 0.342(0.1725/0.5044)下降到 1994 年的 0.169(0.1171/0.6941)。换言之，白人(相对于黑人)在 1974 年要乘以 0.342 才能有很强的民主党人认同感，而在 1994 年是乘以 0.169。因此，在 1974 年，白人明显比黑人更不可能有很强的民主党人认同感，而在 1994 年，这一种族差异更显著了。

请注意，白人和黑人在有很强的民主党人认同感的比数的比例从 0.342 下降到 0.169。1994 年的种族影响比 20 年前更大。在 1994 年(相对于 1974 年)，白人要乘以比数 0.494 才能有很强的民主党人认同感(表 6.1)。因为白人在 1974 年已经比黑人更不可能有很强的民主党人认同，所以把比数比降低一半则进一步扩大了民主党身份认同的种族差异。

下一步是检验表 6.1 报告的种族差异的统计显著性。虽然有多种方法来检验统计显著性，但最有指导意义的是把民主党身份认同作为因变量进行 logistic 回归。用 logistic 回归来检验种族作用的统计显著性后，下一步很容易加入协变量来估计在控制了其他解释变量后种族的净作用。

表 6.2 给出了 logistic 回归分析的结果。我们在表的标题中用术语“总作用”来强调没有控制变量下的作用。这个模型只是重复了表 6.1 报告的粗差别。这个模型是一个参数变化模型，交互项是种族乘以 1994 年，它允许民主党身份认同的种族差异在 1974 年和 1994 年不同。该模型可以写成如下形式：

$$E(Dem) = \alpha + \gamma D_{94} + \beta \times \text{种族} + \delta \times D_{94} \times \text{种族} \qquad [6.7]$$

此处，种族是一个虚拟变量(1 = 白人)。请注意，这里的种族差异模型和前面讨论的地区差异模型在形式上是一样的。两者都只有 4 个参数：截距(α)和 3 种截距差(γ、β 和 δ)。这种类型的模型是参数变化模型的最基本形式。

表 6.2 报告了 4 个参数的估计。因为读者处理百分数的经验可能比处理比数的更多，表 6.2 首先把普通最小二乘回归分析和 logistic 回归分析的结果并列报告。logistic 回归重复了表 6.1 中报告的比数，而普通最小二乘回归重复了表 6.1 的百分数。我用普通最小二乘回归的结果来确定基本原则。在此重要的是要强调这个练习是作为教学，即展示怎么用最小二乘法来估计总作用模型以重现表 6.1 的粗百分数，而不是说我们认可用普通最小二乘估计来对二分的因变量进行回归。在那样的模型中，logistic 回归是更合适的估计统计显著性的方法。表 6.2 报告了普通最小二乘估计的显著性，只是为了让它们可以与 logistic 回归的估计相比较。

表 6.2 民主党身份认同的种族差异[a]（1974 年对 1994 年）：总作用的参数变化模型的回归结果（$N = 5581$）

自变量	普通最小二乘回归[b]		logistic 回归[c]		
	系数	p	logit	比数比	p
表示 1994 年的虚拟变量	7.4*	0.006	0.319*	1.376	0.04
种族（1 = 白人）	−18.8*	<0.0001	−1.074*	0.342	<0.0001
种族×表示 1994 年的虚拟变量	−11.7*	<0.0001	−0.706*	0.494	<0.0001
截距[d]	33.5*	<0.0001	−0.6841*	0.5045	<0.0001

注：a. 自我认定为"很强的民主党人（认同感）"。
b. 普通最小二乘回归报告的系数可以用来重现表 6.1 的百分数。这些系数可以通过以下过程得到，即把二分变量（1 = 很强的民主党人认同感，0 = 其他）回归到表示 1994 年的虚拟变量、种族和种族×表示 1994 年的虚拟变量，并把得到的系数乘以 100，或者为避免给系数乘以 100，可以把因变量编码为（0，100），而不是（0，1）。
c. 在这里，比数比是通过对 logit 取反对数（以 e 为底）而得到的。以上的比数比可以用来重现表 6.1 报告的比数（参照正文的详细介绍）。
d. 参照组，即 1974 年的黑人的系数（百分数或者 logit）。
* 表示 $p < 0.05$。

表 6.1 的百分数可以由普通最小二乘估计重现，只要给虚拟变量代入合适的值。对 1974 年的黑人，我们有：

$$
\begin{aligned}
E(Dem) &= \alpha + \gamma D_{94} + \beta \times \text{种族} + \delta \times D_{94} \times \text{种族} \\
&= 33.5 + 7.4(0) - 18.8(0) - 11.7(0) \qquad [6.8] \\
&= 33.5
\end{aligned}
$$

对于 1994 年的黑人，最后两项都为 0，$E(Dem) = \alpha + \gamma D_{94} = 33.5 + 7.4 = 40.9$。对于 1974 年的白人，种族 = 1 且 $D_{94} = 0$，所以 $E(Dem) = \alpha + \beta \times \text{种族} = 33.5 - 18.8 = 14.7$。对于 1994 年的白人，种族 = 1 且 $D_{94} = 1$，所以 $E(Dem) = \alpha + \gamma D_{94} + \beta \times \text{种族} + \delta D_{94} \times \text{种族} = 33.5 + 7.4 - 18.8 - 11.7 = 10.4$。

这个练习展示了普通最小二乘回归的一项为人所熟知的规律：当所有的自变量都是虚拟变量时，最小二乘回归预测的是用虚拟变量建模的各个分类的均值。在因变量是二分编码（0，100）的特殊情况下，回归系数可恢复为实际的类别百分数。之所以如此，是因为百分数是以（0，100）编码的变量的均值。例如，如果 5 个受访者中有 2 个人有很强的民主党人认同感，

那么均值是 (100＋100＋0＋0＋0)/5 = 40 = 有很强的民主党人认同感的百分数。

同样，用 logistic 回归到虚拟的自变量上的结果也可用于恢复各类别的比数。首先考虑截距。截距是参照组(1974 年的黑人)的均值。在普通最小二乘法的情况下，截距是 1974 年的黑人中自认为有很强的民主党人认同感的人的百分数(方程 6.8)。在 logistic 回归的情况下，截距是在 1974 年非洲裔美国人有很强的民主党人认同感的比数。因为 33.5%的白人在 1974 年有很强的民主党人认同感，所以比数就是 33.5/(100 − 33.5) = 0.504，与表 6.2 的截距所给出的值一致(在四舍五入范围内)。

现在考虑非洲裔美国人在 1994 年有很强的民主党人认同感的比数。从表 6.1 中，我们知道，比数从 0.504 提高到了 0.694。我们也可以从表 6.2 中发现比数在提高，如下所示：

$$
\begin{aligned}
\text{logit}(Dem) &= \alpha + \gamma D_{94} + \beta \times \text{种族} + \delta \times D_{94} \times \text{种族} \\
&= -0.684 + 0.319 D_{94} - 1.074 \times \text{种族} - 0.706 \times \text{种族} \times D_{94} \\
&= -0.684 + 0.319(1) - 1.074(0) - 0.706(0) \\
&= -0.365
\end{aligned}
\qquad [6.9]
$$

−0.365 是非洲裔美国人在 1994 年有很强的民主党人认同感的 logit(比数的对数)。把 logit 转化为比数，我们对 −0.365 取(以 e 为底的)反对数。logit = −0.365 意味着比数 = $e^{-0.365} = 0.694$。

在方程 6.9 中，我们把(logit 的)冥相加，然后取反对数来得到比数。或者，因为 $e^{x+y} = e^x e^y$，我们可以先取反对数，然后相乘(即 $e^x e^y$)而不是相加，再之后取反对数(e^{x+y})。由方程 6.9 可知，对 1994 年的非洲裔美国人，有很强的民主党人认同感的比数是 $e^{\{-0.684+0.319(1)-1.074(0)-0.706(0)\}} = e^{-0.684} e^{0.319} e^0 e^0 = (0.505)(1.376)(1)(1) = 0.695$(根据定义，$e^0 = 1$)。

把反对数相乘经常是更容易的方法，因为一般的 logistic 回归软件在报告 logit 的同时报告反对数。这些反对数在表 6.2 中标记为“比数比”。任意一组

比数可以通过将适当的比数比相乘而计算得到。例如，表 6.2 中的比数比暗示 1994 年非洲裔美国人的比数为 0.695(0.505×1.376)。对黑人来说，从 1974 年到 1994 年，有很强的民主党人认同感的比数要乘以因子 1.376，即明显提高了(比数从 0.505 提高到 0.695)。

白人的情况则很不同。对他们来说，从 1974 年到 1994 年，有很强的民主党人认同感的比数要乘以因子(1.376)(0.494) = 0.680，即下降了。这从表 6.2 报告的比数比可以得到：

$$1974\text{ 年的白人的比数} = (0.505)(0.342) \quad [6.10]$$

$$1994\text{ 年的白人的比数} = (0.505)(0.342)(1.376)(0.494) \quad [6.11]$$

比较方程 6.10 和方程 6.11，很明显，1994 年的比数是 1974 年的比数乘以(1.376)(0.494)或 0.680。对白人来说，有很强的民主党人认同感的比数从 1974 年到 1994 年以因子 0.680 下降，而对非洲裔美国人来说，则以因子 1.376 提高。因此，白人对黑人的比数的比率从 1974 年到 1994 年要乘以 0.680/1.376 = 0.494。

比数比(0.494)的参数变化在统计上是显著的，显示从 1974 年到 1994 年，政党身份认同的种族差异发生了变化。系数 0.494 表示白人有很强的民主党人认同感的比数，相对于黑人有很强的民主党人认同感的比数，从 1974 年到 1994 年要乘以 0.494。因为白人在 1974 年已经比黑人更不可能有很强的民主党人认同感，所以白人对黑人的比数减半使得有很强的民主党人认同感的种族差异扩大了。我们由此得出结论，在 1974 年到 1994 年间，民主党身份认同的种族差异扩大了，至少从自认为有很强的民主党人认同感来说是这样。[11]

还有什么结论可以从表 6.2 中得出？首先，考虑表示 1994 年的虚拟变量的结果。参照组是 1974 年的黑人，所以表示 1994 年的虚拟变量的系数是比较 1994 年的黑人和 1974 年的黑人。(这个)logit 是正的且统计上显著，意味着黑人在 1994 年比在 1974 年有更强的民主党人认同感。换言之，黑人民主党认同感的 7.4 个百分点的提高(表 6.1)具有统计显著性。此外，考虑表示

种族的虚拟变量的结果。因为参照组是 1974 年的黑人，所以这个 logit 比较的是 1974 年的白人和黑人。因为该 logit 是负的且统计上显著，所以我们得出结论，在 1974 年，白人比黑人更不可能有很强的民主党人认同感。换言之，黑人和白人在 1974 年时的 18.8 个百分点的差异不太可能是抽样误差的结果（$p < 0.0001$）。[12]

回归并没有告诉我们白人的民主党人认同感的下降是否在统计上显著。为了进行检验，我们需要把种族的编码对调（例子如下）。

不同种族和教育水平者的民主党人身份认同

我有意选择政党身份认同的种族差异扩大这个例子，来展示各群体的比数怎样从有分类自变量的参数变化模型的 logit 系数中得到。在关于种族和政党身份认同的实际论文中，很自然，下一步是增加一系列协变量，看对民主党身份认同的种族差异是否能被这些协变量所解释。我推延这一步，先着重考察包含连续变量的参数变化模型的基本规律。以教育为例。控制了刚刚观察到的种族差异后，教育与民主党身份认同的关系在 1994 年和 1974 年有没有不同？为了研究这个问题，我在种族总差别的参数变化模型中加入教育和教育乘以 1994 的交互项。

表 6.3 报告了结果。因为我们从一个所有自变量都是分类变量的模型，跳到了另一个有部分自变量不是分类变量的模型中，截距不能再被解释为参照组的均值。相应的，截距是参照组成员的预期值，这些参照组成员在连续变量上取值为 0。因此，这个模型预测，对 1974 年没受过任何正式教育的非洲裔美国人来说，他们有很强的民主党人认同感的比数的对数为 0.41。如果比数的对数为 0.41，比数则为 $e^{0.41}$，或者说是 1.51，所以概率是 1.51/(1.51 + 1)，或者说 0.60。我们也得出结论，0.60 的概率对 1994 年和 1974 年都适用，因为在表 6.3 中，表示 1994 年的虚拟变量没有达到统计显著性。[13]这个结果跟我们在表 6.2 中的发现不一致，在表 6.2 中，黑人在 1994 年比在 1974 年有更大的民主党人认同感的比数。

表 6.3 民主党身份认同的种族和教育差异[a](1974 年对 1994 年)：参数变化模型的回归结果($N = 5581$)

自变量	logit	比数比	p
表示 1994 年的虚拟变量	−0.14	0.87	0.67
种族(1 = 白人)	−0.94*	0.39	<0.0001
种族×表示 1994 年的虚拟变量	−0.80*	0.45	<0.0001
教育	−0.106*	0.90	<0.0001
教育×表示 1994 年的虚拟变量	0.055*	1.06	0.03
截距	0.41*	1.51	0.04

注：a. 自评为“有很强的民主党人认同感”。
*表示 $p < 0.05$。

在模型中加入教育后，种族的作用不变，也就是说，白人比黑人更不可能有很强的民主党人认同感，且这个差异在增大。在教育水平不变的情况下，白人(相比于非洲裔美国人)在 1974 年要乘以比数 0.39 才能有很强的民主党人认同感，而在 1994 年则是乘以(0.39)(0.45) = 0.18。这些估计跟种族的总作用相似，它在 1974 年为 0.342，在 1994 年为(0.342)(0.494) = 0.169(表 6.1)。即使控制了教育的作用，民主党身份认同的种族差异还是扩大了。

至于教育自身的作用，教育程度更高的人更不可能有很强的民主党人认同感。在 1974 年，教育年限每增加一年，要有很强的民主党人认同感就要乘以比数 0.90。交互项教育×表示 1994 年的虚拟变量在统计上显著($p = 0.03$)，显示教育在 1994 年的作用与 1974 年不同。交互项的 logit 系数是 1974 年和 1994 年斜率的差别，因此 1994 年的预期(logit)斜率是 1974 年的斜率加上 1994 年的增长，即 $-0.106 + 0.055 = -0.051$。因为 −0.051 的反对数为 0.95，所以我们可以说在这个样本里，在 1994 年，每增加一年教育，要有很强的民主党人认同感就要乘以比数 0.95(或者，我们可以由 1974 年教育作用的比数比[0.90]乘以 1994 年教育作用的增长[1.06]而得到比数比 0.95)。

总而言之，从表 6.3 报告的结果，我们可以得出结论说，教育在 1974 年对民主党身份认同有相反的作用($p < 0.0001$)，且教育的作用在 1994 年比在 1974 年弱($p = 0.03$)。然而，从表 6.3 中，我们不清楚教育在 1994 年的作用是否在

统计上显著。我们可以计算出教育在1994年的样本斜率并用logit表示(见上)是−0.051,但是我们缺乏对这个logit的显著性检验。

为了检验这个logit的显著性,我们可以把表示年份的虚拟变量的编码对调,重新估计这个参数变化模型。当表示年份的虚拟变量在1994年编码为0,而在1974年编码为1时,教育的logit系数就代表1994年的教育的预期作用,而不是1974年的作用。这就是我们想要检验显著性的系数(要了解另一种对调编码以检验显著性的方法,参见Firebaugh & Beck, 1994)。

表6.4报告了参数变化模型的结果,其中年份的虚拟变量是1974年编码为1。事实上,教育在1994年的作用在统计上是显著的($p = 0.007$,表6.4)。我们得到的logit系数是−0.052,与表6.3得到的计算结果(−0.106 + 0.055)一致。那么,在1994年,每增加一年教育,要有很强的民主党人认同感就要乘以比数$e^{-0.052}$,或者说是0.95。

表6.4 民主党身份认同的种族和教育差异[a](1974年对1994年):参数变化模型的回归结果(对调了年份虚拟变量的编码,$N = 5581$)

自变量	logit	比数比	p
表示1974年的虚拟变量	0.14	1.15	0.67
种族(1 = 白人)	−1.74*	0.18	< 0.0001
种族×表示1974年的虚拟变量	0.80*	2.22	< 0.0001
教育	−0.052*	0.95	0.007
教育×表示1974年的虚拟变量	−0.055*	0.95	0.03
截距	0.27	1.31	0.30

注:a. 自评为"有很强的民主党人认同感"。
*表示$p < 0.05$。

重复一点,表6.4的重点是要检验教育在1994年的作用的统计显著性。除了统计显著性这个问题,没有必要估计表6.4的模型,因为表6.4的自变量的logit和比数比可以从表6.3的系数中计算得来。然而,我把两个表都放在这里,是因为读者会发现对它们进行比较很有意义。例如,读者会发现,把年份的编码对调后,所有包含年份的logit的符号都对调了,包括年份自身、种族×年份和教育×年份。此外,对调了年份的编码后,也对调了模型中种族和教育的叠

加作用的参照组——种族和教育的作用现在指的是 1994 年的种族和教育的作用,而不是 1974 年的。举例来说,种族的比数比是 0.18(表 6.4),是 1994 年的比数比,而不是 1974 年的比数比。

我在此处加入表 6.4,因为它也可以强调一点,即不要把直接检验作用的变化(正如在参数变化模型中所用到的)与对个别变化进行显著性检验混淆。有经验的研究者会注意到这种情况:在群体 A 中,变量 X 的系数的 p 值小于 0.05,因此在 0.05 的显著性水平上,我们的结论是,X 对群体 A 有影响。在群体 B 中,系数稍小且 p 值大于 0.05,因此我们的结论是,X 对群体 B 没有影响,然而 X 的这两个系数之间的差别在统计上不显著。相反,即使两个系数各自并不显著,这个差别也可能在统计上显著(这种情况可能会出现,例如在一个样本的系数为正,另一个为负的情况下)。

研究者要从中学到的是,由变量作用的统计显著性的变化来推论变量的作用的变化要特别小心。如果 X 在时点 1 的作用在统计上显著,而在时点 2 不显著,那么我们不能推论说,X 的作用显著地变化了。参数变化模型提供了检验变量作用变化的直接方法,研究者如果想得出关于作用变化的结论,应该运用这些直接检验的方法,因为 X 的作用在统计显著性上的变化不能保证 X 的作用发生了显著的变化。

种族对民主党身份认同的净作用

从逻辑上来说,研究种族在政党身份认同上的差异的下一步,是增加协变量来检验种族差异以及种族差异的扩大能否被黑人与白人之间的社会经济差异所解释。为了方便展示,我在表 6.4 报告的参数变化模型中加入了地区、性别和年龄(包括一次项和二次项,即年龄和年龄的平方)。这些协变量(还包括教育,它已经在模型中)在豪特、布鲁克斯和曼扎研究投票给民主党总统候选人的阶级差异时也被使用(Hout, Brooks & Manza, 1995)(关于该研究的详细内容,下一节将提到)。

表 6.5 种族和教育对民主党身份认同的净作用[a]（1974 年对 1994 年）：参数变化模型的回归结果（$N = 5581$）

自变量	logit	比数比	p
表示 1994 年的虚拟变量	−0.09	0.91	0.78
种族（1 = 白人）	−1.04*	0.36	<0.0001
种族×表示 1994 年的虚拟变量	−0.83*	0.44	<0.0001
教育	−0.06*	0.94	0.0002
教育×表示 1994 年的虚拟变量	0.04	1.05	0.07
地区（1 = 南方）	0.07	1.07	0.40
性别（1 = 女性）	−0.03	0.97	0.70
年龄[b]	0.04*	1.04	<0.0001
年龄的平方[b]	−0.0002	1.00	0.11
截距	−0.91*	0.40	0.0004

注：a. 自评为"有很强的民主党人认同感"。

b. 年龄的测量为受访者的年龄减去 16。

* 表示 $p < 0.05$。

表 6.6 种族和教育对民主党身份认同的净作用[a]（1974 年对 1994 年）：参数变化模型的回归结果（对调了年份虚拟变量的编码，$N = 5581$）

自变量	logit	比数比	p
表示 1994 年的虚拟变量	0.09	1.09	0.78
种族（1 = 白人）	−1.87*	0.16	< 0.0001
种族×表示 1994 年的虚拟变量	0.83*	2.29	< 0.0001
教育	−0.02	0.98	0.40
教育×表示 1994 年的虚拟变量	−0.04	0.96	0.07
地区（1 = 南方）	0.07	1.07	0.40
性别（1 = 女性）	−0.03	0.97	0.70
年龄[b]	0.04*	1.04	< 0.0001
年龄的平方[b]	−0.0002	1.00	0.11
截距	−1.00*	0.37	0.0008

注：a. 自评为"有很强的民主党人认同感"。

b. 年龄的测量为受访者的年龄减去 16。

* 表示 $p < 0.05$。

表 6.5 和表 6.6 报告了加入协变量后的结果（表 6.5 中年份是 1994 年编码为 1，而表 6.6 中是 1974 年编码为 1）。简言之，种族的净作用和总作用一样。

加入协变量几乎不影响种族以及种族×1994年的系数。在美国，民主党身份认同的种族差异不能被种族在教育、地区、性别或年龄上的差异所解释。

参数变化模型的另一个例子：阶级和民主党身份认同

豪特、布鲁克斯和曼扎最近的研究展示了这样的证据，即“战后时期美国总统选举中阶级与投票行为的历史性重组”(Hout, Brooks & Manza, 1995: 805)。具体来说，从全国选举研究的数据中，他们发现了中产阶级的投票两极分化：专业人员和技术人员对民主党候选人的支持增加，但企业管理人员和销售人员对其支持减少。

只要把上面的种族的参数变化模型中的种族变量替换为对阶级的合适测量，我们就可以得到一个现成的检验方法，来检验中产阶级两极分化的论点是否被美国综合社会调查的数据中民主党身份认同的结果所支持。阶级和政党身份认同的例子在这里尤为合适，因为它显示了在自变量的取值为多分类时，如何运用参数变化模型。在这里，问题涉及的两个类别——专业人员/技术人员对企业管理人员/销售人员——内嵌于一个包含其他职业类别的阶级结构。在变量的取值是多分类时，我们的兴趣专注于其中两个类别在政党身份认同上的变化。

对一个多分变量进行“有针对性的比较”有两种方法。第一种方法是把分析局限于我们感兴趣的类别之间的比较。例如，在检验中产阶级两极分化的假设时，我们可以只分析中产阶级样本。第二种方法是纳入整个样本，把所有其他职业作为剩余分类加入研究。这时，所有除企业管理人员、销售人员、专业人员和技术人员以外的其他职业都成了一个“其他职业”的虚拟变量。为了方便教学，我使用这两种方法来分析中产阶级两极分化的假设。

表6.7报告了3个职业类别(企业管理人员/销售人员、专业人员/技术人员和“其他职业”)在民主党身份认同上的总差别。表格报告了各群体自评为有很强的民主党人认同感的比数，比数由表中报告的百分数计算而来。从比数中，我们很容易计算出我们感兴趣的类别的比数变化(在这里，专业人员/技术人员

相比于企业管理人员/销售人员用“专业人员”相比于“企业管理人员”来缩写)。由此,比数的变化可以与用 logistic 回归分析的总作用模型中合适的参数变化项的比数比相比较。假设在这两个模型中,变量的编码是一致的,那么它们的估计也应该是一致的(在四舍五入范围内)。如果它们不一致,我们就需要找出出错的地方。

表 6.7　民主党身份认同的阶级差异(1974 年对 1994 年):百分数和比数[a]

	百分比		比　数	
	1974	1994	1974	1994
中产阶级($n=1840$)	10.38	12.54	0.116	0.143
企业管理人员和销售人员	11.35	10.25	0.128	0.114
专业人员和技术人员	9.44	15.01	0.104	0.177
其他职业($n=3353$)	19.97	15.72	0.250	0.187
相关的比数比				
1994 年企业管理人员的比数:1974 年企业管理人员的比数			0.891[b]	
专业人员的比数:企业管理人员的比数(1974 年)			0.813[b]	
专业人员的比数:企业管理人员的比数(1994 年)			1.553	
其他职业的比数:企业管理人员的比数(1974 年)			1.953[b]	
其他职业的比数:企业管理人员的比数(1994 年)			1.640	
作用的变化、专业人员:企业管理人员[c]			1.91[b]	
作用的变化、其他职业:企业管理人员[d]			0.84[b]	

注:a. 自评为“有很强的民主党人认同感”。样本量是 5193。样本包括所有在职业、政党身份、性别、年龄和地区上没有缺失值的受访者。

b. 比值在表 6.8 中是 logistic 回归的系数。

c. 专业人员的比数:企业管理人员的比数,1994 年的比值除以 1974 年的比值(1.553/0.813)。

d. 1.640/1.953。

研究的粗略结果似乎支持了中产阶级两极分化的假设。把中产阶级作为整体,有很强的民主党人认同感的比数增加了,从 1974 年的 0.116 提高到了 1994 年的 0.143。这个总体的增加掩盖了中产阶级内部的有趣差别:企业管理人员有很强的民主党人认同感的比数保持不变或者说稍微下降了,然而专业人员的比数则明显提高了。

因此,两个群体在他们对民主党认同感的强度上对调了位置。在 1974 年,企业管理人员更可能有很强的民主党人认同感;在 1994 年,专业人员则更可能

有很强的民主党人认同感。

然而，我们不知道这些样本差异在统计上是否显著。为了检验统计显著性，我们来看 logistic 回归的结果（表 6.8）。跟之前一样，因变量是自评为有很强的民主党认同感的比数的对数。用"针对性比较"的方法，样本限制为职业类别为企业管理人员和专业人员的人。在这种情况下，样本量为 1840，且有 3 个自变量：表示 1994 年的虚拟变量，职业如果是专业人员或技术人员，其虚拟变量（"专业人员"）编码为 1，如果是企业管理人员或销售人员则编码为 0，以及参数变化项，专业人员×1994。用不限制的方法，所有职业都包括了，所以样本量更大（$N = 5193$），这时需要加入两个自变量，一个表示"其他职业"的虚拟变量（如果职业不是企业管理人员或专业人员，该变量编码为 1）和一个参数变化项，即其他职业 × 1994。

表 6.8　中产阶级的民主党身份认同[a]（1974 年对 1994 年）：总作用的参数变化模型的回归结果

自变量	针对性样本[b]（仅中产阶级）			完整样本[c]（所有职业）		
	logit	比数比	p	logit	比数比	p
表示 1994 年的虚拟变量	−0.11	0.89	0.60	−0.11	0.89	0.60
表示专业人员的虚拟变量	−0.21	0.81	0.39	−0.21	0.81	0.39
专业人员 × 1994 年	0.64*	1.90	0.03	0.64*	1.90	0.03
表示其他职业的虚拟变量[d]	—	—	—	0.67*	1.95	<0.0001
其他职业 × 1994 年	—	—	—	−0.18	0.84	0.45
截距	−2.06*	0.128	<0.0001	−2.06*	0.128	<0.0001

注：a. 自评为"有很强的民主党人认同感"。
b. $N = 1840$。包括专业人员和技术人员（在表示专业人员的虚拟变量中编码为 1）及企业管理人员和销售人员。
c. $N = 5193$。
d. 包括所有不被归类为专业人员、技术人员、企业管理人员或销售人员的劳动者。
* 表示 $p < 0.05$。

分析结果支持了中产阶级两极分化的假设。参数变化项专业人员 × 1994 年在这里是重点。专业人员 × 1994 年的比数比是 1.90（$p = 0.03$，表 6.8）。换言之，从 1974 年到 1994 年，专业人员有很强的民主党人认同感的比数相对于企业管理人员的比数的比率，以因子 1.90 发生变化。从表 6.7 来检验数值，专业

人员对企业管理人员的比数比在 1974 年是 0.813，在 1994 年是 1.553，因此从 1974 年到 1994 年，专业人员对企业管理人员的比率要乘以 1.553/0.813，或者说是 1.91（在四舍五入误差内）。

表 6.8 也表明，在 1974 年，专业人员和企业管理人员在民主党身份认同上没有差别。在 1974 年，专业人员和企业管理人员的差别反映在表示专业人员的虚拟变量的系数上：比数比是 0.81，在统计上不显著。比数比的变化在统计上显著（专业人员 ×1994 年）。如前面所解释的，这并不意味着在 1994 年，专业人员和企业管理人员之间的差别在统计上显著，所以我通过对调年份的编码来检验 1994 年的差别的显著性（不在此展示）。实际上在样本中，这个差别是显著的（比数比 = 1.55，$p = 0.02$）。

受限的和不受限的样本在这里得出了同样的结果。这个相同并不是偶然。虽然跟另一种方法相比，受限的样本使用了数据的子样本，但两种方法都是在估计相同的比数，也就是说，都是在估计专业人员和企业管理人员在 1974 年和 1994 年的民主党身份认同的比数。换言之，两种方法都是被设计来重现表 6.7 的相关比数的（在四舍五入误差内），所以得出了相同的结果。

既然针对性样本给出了一样的结果，为什么我们要用整个样本？有两个原因。首先，整个样本给出了中产阶级以外的人在民主党身份认同上的趋势信息，因此我们可以把中产阶级两极分化的发现放在一个更大的背景下。回忆一下，专业人员 ×1994 年是检验相对企业管理人员来说，专业人员在 1974 年至 1994 年间在民主党身份认同上比数的变化。同样，其他职业 ×1994 年也是检验相同的比数比的变化，不过对应的是所有其他劳动者（非中产阶级）相对于企业管理人员的比数比。对非中产阶级劳动者相对于企业管理人员来说，比数比的变化在统计上不显著（$p = 0.45$）。而相对于企业管理人员来说，专业人员转向了民主党，然而这个转向看来并不是对民主党的一般性转向的一部分，如果我们把样本局限于专业人员和企业管理人员，我们将不会了解到这些情况。第二，使用整个样本一般可以更好地估计净作用。正如表 6.8 所展示的那样，两种方法都对总作用给出了一样的结果，因为两种方法都重复了表 6.7 中的比数比。然而，加入控制变量后，两种方法的结果一般不同，因为两个样本中对控制变量

作用的估计一般不同。

表 6.9 中产阶级的民主党身份认同[a](1974 年对 1994 年):有一个连续变量的参数变化模型的回归结果

自变量	针对性样本[b](仅中产阶级)			全样本[c](所有职业)		
	logit	比数比	p	logit	比数比	p
表示 1994 年的虚拟变量	−0.16	0.85	0.06	−0.02	0.98	0.93
表示专业人员的虚拟变量	−0.36	0.70	0.06	0.02	1.02	0.94
专业人员 × 1994 年	0.74*	2.10	0.002	0.60*	1.81	0.049
表示其他职业的虚拟变量	—	—	—	0.49*	1.63	0.005
其他职业 × 1994 年	—	—	—	−0.15	0.86	0.51
教育	−0.10*	0.90	<0.0001	−0.09*	0.91	<0.0001
截距	−0.38*	0.681	0.013	−0.92*	0.40	<0.0001

注:a. 自评为"有很强的民主党人认同感"。
b. $N = 1840$。包括专业人员和技术人员(在表示专业人员的虚拟变量中编码为 1)及企业管理人员和销售人员。
c. $N = 5193$。
d. 包括所有不被归类为专业人员、技术人员、企业管理人员或销售人员的劳动者。
* 表示 $p < 0.05$。

为了说明,我把教育加入中产阶级两极分化的模型中(表 6.9)。在查看表 6.9 时,记住在第一个回归中,对教育作用的估计是基于中产阶级受访者,在第二个回归中是基于整个样本。如果在两个样本中,教育的斜率完全一样,将会是极为罕见的情况。当教育作用的估计值不同时,我们预期其他解释变量作用的估计值也会不同(在这里确实不同),因为其他变量与教育相关。

总结与扩展

第 6 章展示了一个用于在重复调查中研究解释变量作用的变化的模型。这个模型的决定性特征是它有交互项,其形式是 X 乘以 D_{YR},在这里,X 是一个解释变量(分类变量或者连续变量),D_{YR} 是一个表示调查的虚拟变量。这个模型简单,然而灵活。我们用美国综合社会调查的例子,以一定的详细程度展示了这个模型的灵活性,并且解答了在使用模型中遇到的实际问题。

重要之处在于，这个模型是用于记录变化，而不是解释变化，还需要进一步的分析——或者用不同的模型——来解释观察到的变化。其中一种很有潜力的方法是把个人层次的参数作为随时间变化的宏观层次特征的函数（DiPrete & Grusky, 1990; Firebaugh & Haynie，待出）。这个方法模仿了研究学校作用的多层次方法（Bryk & Raudenbush, 1992），只是在此，背景是时间，而不是学校。

用多层次方法来研究随时间变化的作用，首先是把个人层次的参数本身作为宏观层次变量的函数和一个随机扰动项。由于在宏观层次方程中有随机扰动项，所以多层次模型是一个随机效应模型，且普通最小二乘法不适用。前面描述的交互模型是一个固定效应模型，因此（取决于因变量的特征）它可以用普通最小二乘法或 logistic 回归来估计。与之不同的是，多层次模型需要的估计方法超出了本书的范围（DiPrete & Grusky, 1990）。

第 7 章 | 总结:分析社会变迁

正如在序言中提到的,我曾想把本书命名为《如何用重复调查来分析社会变迁》。本书的核心是介绍 4 种方法来研究社会变迁。这些方法之所以不同,是因为它们提出了关于变迁的不同问题。

第一种方法是趋势分析。趋势分析问的是在某些群体中,Y 的平均值是否随时间变化。因此,在趋势分析中,$E(Y)$表达为时间的函数。因为这些是群体的趋势,而不是个人的,所以这里介绍的趋势分析是宏观层次的(对个人的趋势分析要求有固定样本数据)。很常见的情况是,我们的兴趣在于不同群体的趋势是趋同还是趋异。第 3 章介绍和演示了如何探索这个问题。

第二种方法是对趋势进行近似分解。此时,关注点是社会变迁的近似原因,即有多少是由于个人的净变化带来的,有多少是由人口更替带来的?第 4 章介绍了两种分解方法——一种基于线性回归,另一种基于代数。

第三种方法是用其他变量自身和作用的变化来分解一个变量的变化。第 5 章介绍了按照这样的方法来分解总变化的一般模型。参数变化模型基于回归标准化的分解方程,但将之应用于对变迁的分解上,而不是群体差异的分解上。

第四种方法关注于个人层次上变量作用的变化。第 6 章介绍了一个模型——参数变化模型——来研究个人层次的关系是否随时间变化。参数变化模型在表面上与研究趋势的趋同/趋异的模型相似(第 3 章),因为两个模型都有与时间的交互项,但相似点仅限于此。因为趋势是 Y 的预期值随时间变化而发生的变化,时间在趋势分析中被看做连续的,且 $E(Y)$ 表达为时间的函数。相反,参数变化模型把 $E(Y)$看做在离散时点上 X 的函数,在这里,X 对 $E(Y)$的作用是时间的函数。

因为时间本身并不是一个因果变量，趋势分析只是描述性的。此外，趋势分析通常只纳入有限的变量。例如，表 3.1 报告的趋势分析只纳入了 3 个变量（但是使用了从 1973 年到 1993 年的美国综合调查）。在最简单的情形下，趋势分析是双变量的：变量 Y 和时间。但是，很常见的情况是，趋势分析的目的是研究不同人口子群体的趋势是否不同。在这种情况下，至少有 3 个变量：因变量 Y、时间和一个表示子群体的变量。在某些情况下，子群体反映了多个变量类别的分类。例如，菲尔鲍和哈雷检验了工作满意度在不同种族一性别群体中的情况（白人男性、白人妇女等等）(Firebaugh & Harley, 1995)。我们很少对多于 2 个或 3 个维度的交互分类的群体检验趋势，因为这会导致需要检验的趋势的剧增，以及伴随的各趋势的样本量减少。因此，趋势模型倾向于简洁。

尽管趋势分析只需要少量变量，但它要求那些变量有多次测量。多次测量是这里的要点。变迁分解和参数变化分析都只要求两个时点（但通常纳入更多的变量），因此只重复了一次或两次的调查可用于这里介绍的变迁分解和参数变化模型，但不能用于严格的趋势分析。谈及近似分解，世代更替的作用最适用于由多次测量来估计——在那种情况下，可以使用线性分解法——然而估计世代更替在两次调查间对变迁的贡献也是可行的（正如前面的多个例子所展示的那样）。除此之外，近似分解对数据的要求并不高。区分世代更替的贡献和个人变化的贡献，只需要测量 3 个变量：因变量、出生年份和调查年份。

这 4 种方法——趋势分析、近似分解、变迁分解和参数变化模型——都利用了重复调查的重复性。这些模型提供了简单但有用的工具来研究社会变迁。但工具的价值最终要在使用中被证明，这些工具对研究社会变迁的价值也只有在社会科学家使用它们的时候才能实现。

注释

[1] 重复调查的一个优点是，它们可以合并起来以获得足够的样本量，使我们可以研究通常难以进行分析的人口中的一些亚群体。例如，使用美国综合社会调查，我们可以研究天主教徒或离婚者，甚至是离婚的天主教徒。然而，因为需要时间来积累样本，这样的研究一般必须假定（自变量的）作用是随时间不变的。

[2] 在某些情况下，世代的差别可能来自他们样本的相对大小。例如，一个相对较小世代的成员可能由于学校中班级的规模更小及进入就业市场的竞争更小而获益，此时对较大的世代来说，情况则相反。我们不着重于这种世代效应，因为与其他形式的世代效应相比，在实证上并不难将基于世代大小的作用与年龄和时期的作用区分开来。我们可以很简单地将世代大小作为一个变量加入测量了年龄和时期的模型。将其他形式的世代效应与年龄和时期作用区分开来要难得多。

[3] 用美国综合社会调查的变量 OVERSAMP 加权，以便把在 1982 年和 1987 年以更大的权重抽取黑人的样本调整过来。

[4] 因为这些估计假定了线性趋势，所以它们一般与基于首次调查本身的均值进行的估计不同。

[5] 重要的是，不要把世代间的斜率 β_2 与世代效应相混淆，因为（在此重复）β_2 可以反映年龄作用，也可以反映世代效应。

[6] 参见 Firebaugh，1992。它讨论了为什么代数分解法在分解间隔较大时可能高估世代更替的作用。正如在这篇文章中讨论的，代数分解法用在间隔时间较短的调查中效果最好。要注意，将世代更替的部分进一步分解（例如，分解为源于进入和退出世代而导致的部分和源于在持续存在的世代里死亡率差异而导致的部分）并不那么直接。

[7] 美国综合社会调查抽取的是家庭而不是个人，因此单身的成年人被抽取的权重更大。因为婚姻状况可能与性别角色态度相关，所以我用人口现状调查的婚姻状况数据对美国综合社会调查数据进行了加权。

[8] 所有问题都编码为 0 或 1。0 = 传统的回应，因此均值的增加表示对传统态度支持的减弱。这个减弱不太可能主要来自对社会期望之回应的感知的变化。正如以下所揭示的，这种减弱很大程度上来自不同世代在性别角色态度上的不同，而不太可能是不同世代受社会期望的影响不同。

[9] “1972 年至 1988 年活着的一般成年人的变化”指的是那些出生于 1899 年至 1954 年的人，因为美国综合社会调查的受访者年龄介于 18 岁到 89 岁。那么，为了估计一般的活着的人的变化，人们可能鲁莽地把那些世代作为一个组分开，用 1988 年的均值减去 1972 年的均值。然而，这个捷径无法估计活着的人的平均变化，因为它对 1972 年而非 1988 年，用了世代的比重（p_i）来计算 1972 年活着的人的均值。很显然，除非年长的和年轻的世代有一样的死亡率，否则 1972 年的 p_i 与 1988 年的会不同。

[10] 特谢拉指出，在他的分析中加入参数变化项后，时间虚拟变量的系数（即截距）变得“无法解释”（Teixeira，1987：77），所以他只汇报了叠加分解法的结果。无法解释的截距是前面描述的起点依赖问题的症结，起点依赖很可能是特谢拉的参数变化分析中遇到的问题。

[11] 艾布拉姆森与同事分别汇报了 1952 年到 1992 年全国选举研究数据中白人和黑人的政党认同。有趣的是，全国选举研究在 1974 年和 1992 年的数据在用“有很强的民主党人认同感”时，没有显示出种族差异的扩大。1974 年相对于 1992 年，白人和黑人“有很强的民主党人认同感的百分比”的变化都几乎是 0。研究美国综合社会调查和全国选举研究数据在这点上的区别不属于这里的分析的范畴。我分析美国综合社会调查数据中的种族差异，而全国选举研究数据中的种族

差异问题留给其他人去研究。

[12] 请注意，表 6.2 汇报的系数是基于用虚拟变量编码的自变量。这点值得强调，因为 logit 分析经常对分类变量使用效果编码，而不是虚拟编码。举例来说，不是把种族编码为 1 代表白人，0 代表黑人，种族可以编码为 1 代表白人，－1 代表黑人。效果编码可能产生一样的结果，但结果也可能看起来不同，因为自变量被赋予了不同的值。对于二分变量（正如表 6.2 那样），把用效果编码的 logit 乘以 2 就可将其转化为用虚拟编码得到的 logit。

[13] 在表 6.3 里表示 1994 年的虚拟变量的样本 logit 是－0.14。零假设为，总体的 logit 是 0。当 logit 为 0 时，比数比为 1（因为根据定义，$e^0 = 1$）。因此，与之等同的零假设则是比数比为 1。很显然，比数比为 1 意味着没有作用，因为乘以 1 没有任何作用。例如，在表 6.3 的模型中，在 1994 年，一个没有受过任何教育的非洲裔美国人有很强的民主党人认同感的比数的估计值为 1.51（1974 年的比数）乘以表示 1994 年的虚拟变量的比数比。因为我们无法拒绝零假设，即 1994 年的比数比为 1，所以有 1.51(1) = 1.51。

参考文献

Abramson, P., & Aldrich, J. H. (1982). "The decline of electoral participation in America." *American Political Science Review*, *76*, 502—521.

Abramson, P., Aldrich, J. H., & Rohde, D. W. (1994). *Change and Continuity in the 1992 Elections*. Washington, DC: Congressional Quarterly.

Agresti, A., & Finlay, B. (1986). *Statistical Methods for the Social Sciences* (2nd ed.). San Francisco: Dellen.

Alexander, J. C., Geisen, B., Munch, R., & Smelser, N. J. (Eds.) (1987). *The Macro-Micro Link*. Berkeley: University of California Press.

Bennett, S., & Bennett, L. (1987). "Political participation." *Annual Review of Political Science*, *2*, 157—204.

Brody, R. A. (1978). "The puzzle of political participation in America." In A. King (Ed.), *The New American Political System* (pp. 287—324). Washington, DC: American Enterprise Institute.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Campbell, A., Converse, P., Miller, W., & STOKES, D. (1960). *The American Voter*. New York: Wiley.

Clogg, C. C., & Eliason, S. R. (1986). "On regression standardization for moments." *Sociological Methods and Research*, *14*, 423—446.

Coleman, J. S. (1986). "Social theory, social research, and a theory of action." *American Journal of Sociology*, *91*, 1309—1335.

Comte, A. (1974). *The Positive Philosophy, Book Ⅵ: Social Physics* (Translated by Harriet Martineau). New York: AMS. (Original work published 1839)

Das Gupta, P. (1978). "A general method of decomposing a difference between two rates into several components." *Demography*, *15*, 99—112.

Das Gupta, P. (1993). *Standardization and Decomposition of Rates: A User's Manual*. Washington, DC: Government Printing Office.

Davis, J. A. (1992). "Changeable weather in a cooling climate atop the liberal plateau: Conversion and replacement in forty-two General Social Survey items, 1972—1989." *Public Opinion Quarterly*, *56*, 261—306.

Davis, J. A., & Smith, T. W. (1992). *The NORC General Social Survey*. Newbury Park, CA: Sage.

Davis, J. A., & Smith, T. W. (1994). *General Social Surveys, 1972—1994 [MRDF]* (NORC ed.). Chicago: NORC.

Day, C. L. (1990). *What Older Americans Think*. Princeton, NJ: Princeton University Press.

Diprete, T. A., & Grusky, D. B. (1990). "The multilevel analysis of trends with repeated cross-sectional data." In C. C. Clogg (Ed.), *Sociological Methodology 1990* (pp. 337—368). Oxford,

UK: Basil Blackwell.

Donza, M., Duncan, G. J., Corcoran, M., & Groskind, F. (1988). "The guns of autumn? Age differences in support for income transfers to the young and the old." *Public Opinion Quarterly*, *52*, 441—466.

Duncan, G., & Kalton, G. (1987). "Issues of design and analysis of surveys across time." *International Statistical Review*, *55*, 97—117.

Durkheim, E. (1938). *The Rules of Sociological Method*. Glencoe, IL: Free Press. (Original work published 1895)

Easterlin, R. (1980). *Birth and Fortune: The Impact of Numbers on Personal Weyare*. New York: Basic Books.

Elder, G. H., Jr. (1974). *Children of the Great Depression*. Chicago: University of Chicago Press.

Finkel, S. E. (1995). *Causal Analysis With Panel Data*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—105. Thousand Oaks, CA: Sage.

Firebaugh, G. (1989). "Methods for estimating cohort replacement effects." In C. C. Clogg(Ed.), *Sociological Methodology* 1989(pp. 243—262). Oxford, UK: Basil Blackwell.

Firebaugh, G. (1990). "Replacement effects, cohort and otherwise: Response to Rodgers." In C. C. Clogg(Ed.), *Sociological Methodology 1990*(pp. 439—446). Oxford, UK: Basil Blackwell.

Firebaugh, G. (1992). "Where does social change come from? Estimating the relative contributions of individual change and population turnover." *Population Research and Policy Review*, *11*, 1—20.

Firebaugh, G., & Beck, F. D. (1994). "Does economic growth benefit the masses? Growth, dependence, and welfare in the Third World." *American Sociological Review*, *59*, 631—653.

Firebaugh, G., & Chen, K. (1995). "Vote turnout of Nineteenth Amendment women: The enduring effect of disenfranchisement." *American Journal of Sociology*, *100*, 972—996.

Firebaugh, G., & Davis, K. E. (1988). "Trends in antiblack prejudice 1972—1984: Region and cohort effects." *American Journal of Sociology*, *94*, 251—272.

Firebaugh, G., & Harley, B. (1991). "Trends in U. S. church attendance: Secularization and revival, or merely life-cycle effects?" *Journal for the Scientific Srudy of Religion*, *30*, 487—500.

Firebaugh. G., & Harley, B. (1995). "Trends in job satisfaction in the United States by race, gender, and type of occupation." In R. L. Simpson and I. H. Simpson(Eds.), *Research in the Sociology of Work*, *Vol. 5*: *The Meaning of Work*(pp. 87—104). Greenwich, CT: JAI.

Firebaugh, G., & Haynie, D. L. (in press) "Using repeated surveys to study aging and social change." In M. A. Hardy(Ed.), *Conceptual and Methodological Issues in the Study of Aging and Social Change*. Thousand Oaks, CA: Sage.

Fox, M. F., & Firebaugh, G. (1992). "Confidence in science: The gender gap." *Social Science Quarterly*, *73*, 101—113.

Gallup, G., & Newport, F. (1991). "For first time, more Americans approve of interracial marriage than disapprove." *Gallup Poll Monthly* (August), 60—61.

Glenn, N. (1976). "Cohort analysts' futile quest: Statistical attempts to separate age, period, and cohort effects." *American Sociological Review*, *41*, 900—904.

Glenn, N. (1977). *Cohort Analysis*. Sage University Papers series on Quantitative Applications in the Social Sciences, 07—005. Beverly Hills, CA: Sage.

Glenn, N. (1980). "Values, attitudes, and beliefs." In O. G. Brim, Jr., and J. Kagan(Eds.), *Constancy and Change in Human Development*(pp. 596—640). Cambridge, MA: Harvard University Press.

Hagenaars, J. A. (1990). *Categorical Longitudinal Analysis: Loglinear Panel, Trends, and Cohort Analysis*. Newbury Park, CA: Sage.

Hout, M., Brooks, C., & Manza, J. (1995, December). "The Democratic class struggle in the United States, 1948—1992." *American Sociological Review*, *60*, 805—828.

Johnston, D. F. (Ed.)(1981). *Measurement of Subjective Phenomena*. Washington, DC: Government Printing Office.

Jones, F. L., & Kelley, J. (1984). "Decomposing differences between groups: A cautionary note on measuring discrimination." *Sociological Methods and Research*, *12*, 323—343.

Kalton, G. (1983). *Introduction to Survey Sampling*. Sage University Papers series on Quantitative Applications in the Social Sciences, 07—035. Beverly Hills, CA: Sage.

Kendall, M. G. (1973). *Time-Series*. New York: Hafner Press.

Kiecolt, K. J., & Nathan, L. E. (1985). *Secondary Analysis of Survey Data*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—053. Beverly Hills, CA: Sage.

Kish, L. (1983). "Data collection for details over space and time." In T. Wright(Ed.), *Statistical Methods and the Improvement of Data Quality*(pp. 73—84). New York: Academic Press.

Kish, L. (1986). "Timing of surveys for public policy." *Australian Journal of Statistics*, *28*, 1—12.

Kitagawa, E. (1955). "Components of a difference between two rates." *Journal of the American Statistical Association*, *30*, 1168—1194.

Kleppner, P. (1982). *Who Voted? The Dynamics of Electoral Turnout, 1870—1980*. New York: Praeger.

Kotlikoff, L. J. (1992)*Generational Accounting*. New York: Free Press.

Lesthaege, R., & Surkyn, J. (1988). "Cultural dynamics and economic theories of fertility change." *Population and Development Review*, *14*, 1—45.

Long, J. S., & Miethe, T. D. (1988). "The statistical comparison of groups." In J. Scott Long (Ed.), *Common Problems/Proper Solutions*(pp. 108—131). Newbury Park, CA: Sage.

Longman, P. (1987). *Born to Pay: The New Politics of Aging in the United States*. Boston: Houghton Mifflin.

Mannheim, K. (1952). "The problem of generations." In *Essays on the Sociology of Knowledge* (pp. 276—322). London: Routledge and Kegan Paul. (Original work published 1927)

Markus, G. (1979). *Analyzing Panel Data*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—018. Beverly Hills, CA: Sage.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, K. W. (1973). "Some methodo-

logical issues in cohort analysis of archival data." *American Sociological Review*, *38*, 242—258.

Menard, S. (1991). *Longitudinal Research*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—076. Newbury Park, CA: Sage.

Merriam, C. E., & Gosnell, H. F. (1924). *Non-Voting: Causes and Methods of Control*. Chicago: University of Chicago Press.

Norpoth, H. (1987). "Under way and here to stay: Party realignment in the 1980s?" *Public Opinion Quarterly*, *51*, 376—391.

Paset, P. S., & Taylor, R. D. (1991). "Black and white women's attitudes toward interracial marriage." *Psychological Reports*, *69*, 753—754.

Rasinski, K. A. (1988). "The effect of question wording on public support for government spending." Chicago: NORC.

Rodgers, W. (1990). "Interpretingthe components of time trends." In C. C. Clogg(Ed.), *Sociological Methodology 1990* (pp. 421—438). Oxford, UK: Basil Blackwell.

Rosenbaum, W. A., & Button, J. W. (1992). "Perceptions of intergenerational conflict: The politics of young vs. old in Florida." *Journal of Aging Studies*, *6*, 385—396.

Ryder, N. B. (1965). "The cohort as a concept in the study of social change." *American Sociological Review*, *30*, 843—861.

Sayrs, L. W. (1989). *Pooled Time Series Analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—070. Newbury Park, CA: Sage.

Smith. H. L., Morgan, S. P., & Koropeckyj-Cox, T. (1996) "A decomposition of trends in the nonmarital fertility ratios of blacks and whites in the United States, 1960—1992." *Demography*, *33*, 141—151.

Smith, T. W. (1987). "That which we call welfare by any other name would smell sweeter: An analysis of the impact of question wording on response patterns." *Public Opinion Quarterly*, *51*, 75—83.

Smith, T. W. (1988). *Timely Artifacts: A Review of Measurement Variation in the 1972—1988 GSS Data* (Methodological Report No. 56). Chicago: NORC.

Smith, T. W. (1992). "Some thoughts on the nature of context effects." In N. Schwarz & S. Sudman(Eds.), *Context Effects in Social and Psychological Research* (pp.163—184). New York: Springer-Verlag.

Smith, T. W. (1993). *Is there real opinion change?* (Social Change Report No. 36). Chicago: NORC.

Smith, T. W., Arnold, B. J., & Wesely, J. K. (1995). *Annotated Bibliography of Papers Using the General Social Surveys* (10th ed.). Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Sobel, M. E. (1983). "Some large-sample standard errors for components of a mean difference under a linear model." In S. Leinhardt (Ed.), *Sociological Methodology 1983—1984* (pp. 169—193). San Francisco: Jossey-Bass.

Stephenson, C. B. (1978). *Weighting the General Social Surveys for Bias Relating to Household Size* (Methodological Report No. 3). Chicago: NORC.

Teixeira, R. A. (1987). *Why Americans Don't Vote: Turnout Decline in the United States 1960—1984*. New York: Greenwood.

Teixeira, R. A. (1992). *The Disappearing American Voter*. Washington, DC: Brookings Institution.

Verba, S., & Nie, N. (1972). *Participation in America*. New York: Harper & Row.

Vinovskis, M. A. (1993). "An historical perspective on support for schooling by different age cohorts." In V. Bengtson & W. A. Achenbaum(Eds.), *The Changing Contract Across Generations*(pp. 45—64). Hawthorne, NY: Aldine de Gruyter.

Wilmoth, J. R. (1990). "Variation in vital rates by age, period, and cohort." In C. C. Clogg(Ed.), *Sociological Methodology 1990*(pp. 295—336). Oxford, UK: Basil Blackwell.

Wolfinger, R. E., & Rosenstone, S. J. (1980). *Who Votes*? New Haven, CT: Yale University Press.

译名对照表

a panel study design	固定样本研究设计
aggregate change	总变化
aggregate trends	总趋势
algebraic decomposition	代数分解
British Social Attitudes Survey	英国社会态度调查
cohort effects	世代效应
coincident trends	一致趋势
conditional mean	条件均值
Consumer Surveys	消费者调查
converging trends	趋同趋势
crossed trends	交叉趋势
cumulated data	累积数据
Current Population Survey	人口现状调查
data setup	数据组织
diverging trends	趋异趋势
dummy coding	虚拟编码
effect coding	效果编码
erratic fluctuations	偶然的变动
gross change	总变化
linear decomposition	线性分解
moving averages	移动平均值
National Election Study	全国选举研究
National Health Interview Survey	全国健康访问调查
net change	净变化
net effect	净作用
parallel trends	平行趋势
period effects	时期作用
periodic surveys	定期调查
pooled data	合并数据
population turnover	人口更替
proximate decomposition	近似分解
raw differences	粗差别
repeated cross-sectional design	重复截面设计
repeated surveys	重复调查
rotating panel survey	轮换固定样本调查

smoothing trends	平滑趋势
split panel survey	分批固定样本调查
Survey on Income and Program Participation	收入与项目参与调查
the changing-parameter model	参数变化模型
unit-specific terms	单位层次的特点
voting-age population	适龄投票人口

多元时间序列模型

作 者 简 介

帕特里克 · T. 布兰特(Patrick T. Brandt)

美国得克萨斯大学达拉斯分校经济、政治和政策科学学院副教授。曾在北得克萨斯大学和印第安纳大学任教,并曾是哈佛—麻省理工学院数据中心的一名工作人员。主要教授社会科学研究方法和社会统计学方面的课程,当前的研究集中在发展和运用时间序列模型来分析政治制度、政治经济以及国际关系。

约翰 · T. 威廉姆斯(John T. Williams)

美国加州大学河滨分校政治学系讲座教授。他曾在芝加哥伊利诺伊大学和印第安纳大学布卢明顿分校担任教职。主要研究政治经济和政策分析中的统计方法,曾与人合著《混合的困境:民主、集体行动和超级对手》和《公共政策分析:一个政治经济学方法》。他是政治学领域中运用新的统计分析方法的领军人物,尤其是向量自回归模型、贝叶斯估计以及事件时间序列模型。

译 者 简 介

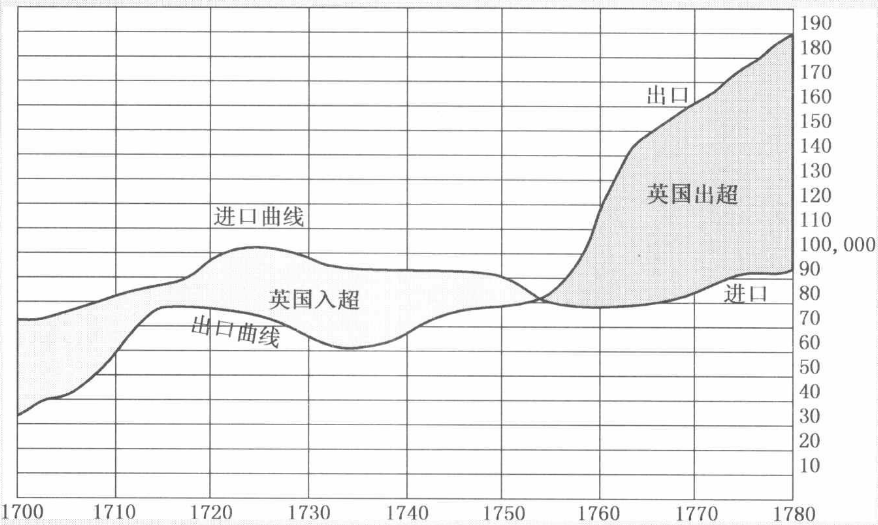
辛济云

2009 年毕业于中国人民大学社会学系，现为香港科技大学社会科学部硕士研究生。研究方向为组织社会学、经济社会学、中小企业融资。

序

社会学家和经济学家长期以来都着迷于时间序列数据的运用。对此类数据的第一个系统的探究来自威廉·普雷菲尔(William Playfair)的《商业和政治图集》,该书发表于220年前,包含43个时间序列图。通过将英国的国债随时间的变化绘制成图,普雷菲尔清楚地发现了1701年后,1730年西班牙战争和1775年美国独立等历史事件对英国国债的影响。

普雷菲尔的图形包含多于一个时间序列的数据。下图的曲线就是进口和出口随时间变化的图形。清楚地证明了两者之间的关系以及两者与时间的关系。这样的图形说明两个时间序列可能不是相互独立的过程。



注:横轴是年,右纵轴的单位是10000英镑。

资料来源:www.unc.edu/~nielsen/soco208/m2/m2033.jpg。

英国对丹麦和挪威的出口以及从丹麦和挪威的进口额(1700—1780年)

尽管时间序列图如此有用，但普雷菲尔的分析依然留下了许多有待进一步回答的问题。是什么导致了进口和出口的增长或下降？进口的数量会对出口的数量产生影响么？或者反之亦然？也许更重要的是，是什么因素打破了进出口之间的平衡？回答这些问题需要对动态的同时过程作出更为合适的分析，布兰特和威廉姆斯的《多元时间序列模型》被视为这方面分析的一个有价值的尝试。

本书作者讨论了4种主要的时间序列数据建模方法：自回归整合移动平均模型、同时方程模型、误差纠正模型和向量自回归模型。他们集中于向量自回归模型的设定、估计和推论，以及格兰杰因果关系检验和通过冲击反应函数来对变量之间动态关系进行评价。同时，本书还提供了两个向量自回归模型的具体运用实例。

廖福挺

第1章 | 前　言

本书的完成历时几年。1999年，我第一次提出写作这本书的想法，那时约翰还在密歇根大学的校际政治与社会研究协会讲授时间序列分析的课程，而我是他的助教。但这一想法一直被搁置，直到约翰在2002年重提此事，当然，我们在这一过程中曾多次讨论。

遗憾的是，正在我们合力写作本书的过程中，约翰于2004年9月去世了。在此之前，本书绝大部分的提纲和写作计划已经完成。约翰对本书的影响甚大，从如何将观点全面展示出来到如何进行多元时间序列分析。在写作过程中，我尽可能忠于我和约翰有关本书的最初想法。

本书包括了多元时间序列分析中的一些高级方法。多元时间序列分析是指含有多于一个（内生）变量的时间顺序数据。举例来说，一个模型认定美国人中民主党人的数量决定了公众对于美国政府实施的政策的支持程度，但反过来想一想，这些政策实际上是被不同政党的参与率所影响的。变量中的一些关系可以被最近的历史因素所解释（过去值），也可能被同时期的因素所解释。所以，我们必须在建立模型时考虑变量之间内生的动态关系。本书致力于描述和推断这种内生的动态关系。

本书在写作时对读者的知识背景做了一些假定。首先，我们认为读者在代数矩阵和方程系统矩阵表达形式方面具有基本的知识。第二，读者应该对普通的线性回归模型有一些了解。最后，读者对单变量时间序列分析应该具有一定的背景知识。

本书的基本写作框架如下：首先对多元时间序列模型的方法论选择和哲学基础做简单的讨论。我们从受到外部因素影响的同时方程模型和单变量时间

序列模型展开讨论。这些讨论的目的是为了阐明不同设定和方程识别假设之间关系。随后，我们讨论了不同的时间序列模型和同时方程模型之间的关系。最后我们介绍了向量自回归模型来代替前两种方法。

第 3 章着眼于同时方程模型和向量自回归模型的具体机制。我们讨论了向量自回归模型的设定问题和一些数学推导。其中包含对滞后项设定、格兰杰因果检验(外生性检验)、冲击反应的动态分析和预测误差方差的分解，以及向量自回归模型和误差纠正模型之间的关系。

第 4 章是向量自回归模型的两个运用实例。第一个例子是分析公众对政府政策的态度和宏观的各政党参与情况之间的关系。第二个例子更为复杂，是讨论有效公司数量的政治和经济原因。在本章中，我们将讨论如何根据数据和理论的要求来对模型进行设定。例子中所用到的数据和软件命令(用 RATS 软件的程序语言写成)可以在以下链接中找到：http://www.utdallas.edu/~pbrandt。这些内容可以帮助读者体验研究的过程是如何进行的。

在附录中，我们讨论了如何选用适合时间序列分析的统计软件。

第 2 章 | 对多元时间序列模型的介绍

很多社会科学数据的问题在实质上涉及多元变量并具有动态性。例如，社会公众对于总统工作表现的满意程度与国家总体的经济状况是什么关系？各国的军费投入是彼此相关的还是外生的？A 国对 B 国所采取的行动是否和 B 国对 A 国所采取的行动相关？各主要政党参加人数在美国人中所占的比重是否和他们对国家政策的支持度相关？税率和有商业背景的政治行动团体的比例是什么关系？在上述每一个例子中，我们都可以写出一个方程，将一个变量作为因变量而其他变量作为自变量。但是，这些例子中的不同变量之间可能存在同时性，因此可能存在第二个方程，其中自变量和因变量之间的关系是相反的。

在上述研究问题中，自变量和因变量都可能是内生的。我们发现，存在一些因素既可以解释各政党参加的人数，又和政府的政策存在动态的和内生性的联系。同样，税率的变化会影响具有公司背景的政治行动团体的游说活动，但其同时也受到这些游说活动的影响。在这两个例子(我们将在第 4 章中详细讨论)中，研究者必须考虑两个(或多个)方程，针对每个方程中的每个变量都列出一个方程，并且让方程中每个变量过去的值和现在的值相互影响。

大多数社会科学家在接受统计训练的初期都学习使用回归方程。但是，单一方程的回归模型忽视了这样一个事实，即面对内生的、动态性的关系时，必须利用多个回归方程。分析者可以继续用单一的回归方程来进行估计并寄希望于估计结果不会产生太大的纰漏，或者，分析者可以选择用计量经济学中发展出来的各种技巧来运用多方程模型进行估计。例如似乎不相关回归、分布滞后自回归(ADL)、转换函数模型。但即使是这些方法，也只考虑了估计中极个别

的问题，例如，序列相关和内生性。分析者必须面对一些更为复杂的问题，例如，数据是按照时间顺序记录的并且需要将时间段作为分析单位。研究者在处理这些问题时，需要估计一个或多个方程并考虑时间序列的动态性。这就需要我们考虑其他分析方法的优点，例如矩阵自回归、误差修正模型以及（动态）因子分析。因为时间序列数据更有价值——换句话说，它比横截面数据包含了更多的信息——所以如何处理多元时间序列问题是十分关键的。

对动态多方程模型的需求来源于社会科学模型中的两个基本事实。首先，变量之间同时相互影响，因此都被认为是内生变量。而在一个多方程系统中，方程的数量经常和内生变量或因变量的数量相同，但也有例外。[1]尽管分析者的理论兴趣仅仅集中在某一个单独的方程上，并且也只对该方程进行估计，但统计和计量经济理论却要求将所有的方程都考虑进去，否则统计推断将会是有偏或者无效的。其次，当我们考虑多个因变量之间的关系时，每个方程所反映出的变量之间单一的、给定的关系必须放在整个系统中加以考虑。要识别变量之间的关系，最合适的做法就是参考所有方程所提供的信息，这就要求在由所有方程所组成的系统中，有足够多的内生变量被正确的估计。估计需要利用整个系统中的内生变量来对变量之间的关系提供尽可能无偏和有效的估计。[2]

在处理这些问题时，我们希望将变量之间的动态关系放在首位。我们希望知道，一个变量的变化如何影响其他变量，但变量之间的关系可能是内生性的。首先，一个变量的变化对另一个变量的影响可能是延后的（所以是前一个变量过去的值对后一个变量当下的值产生影响）。另一种情况是，这种关系的变化来自整个方程系统的变化，例如，我们知道的由冲击和创新带来的系统性变化会导致两个或者多个变量同时发生变化，这是因为对一个变量的冲击可能与对另一个变量的冲击是相关的。

在将理论转化成经验的可估计模型的过程中（是指我们对参数进行估计，然后推断总体的过程），一个中心问题是，我们可能并不知道方程或方程之间的关系结构是否真实地反映出模型。也就是说，假设在一些多元概率密度函数 $f(y|\beta)$中，利用一系列参数 β 来描述我们观测到的数据 y。如果没有另一组参数 β 的值可以产生同样的概率密度，那么这一概率密度函数将会确定唯一一组

的变量关系结构或者一组方程。[3]作为社会科学家，我们并不确定我们写下的方程是否能正确反映具体的模型。这个问题所造成的结果是，许多关于模型及参数解释方面的争论并不是基于模型本身的性质，而是在用来表示模型的方程以及变量之间的关系结构上存在分歧。[4]本书的目的就是列举几种选择方案，帮助社会科学家建立理论并构造时间序列模型。我们将首先强调一些主要的方法，然后描述这些方法的意义。

社会科学理论的建立遵循以下几个步骤：

首先，研究者必须确定理论中涉及的一些主要变量以及变量之间的关系。在这一过程中，研究者需要审视主要的理论（或一系列理论）来指导其经验问题并且确定构造模型过程中的主要变量以及它们之间的关系。即使存在竞争性的理论，这一过程也不存在问题，因为我们假设变量的时间序列相关是可以被测量的。

其次，一旦理论中涉及的主要变量被确定，研究者就开始进入一个关键阶段，即为模型选择合适函数形式以及数学结构。正是在这一过程中，我们会发现同一个理论可以有不同的模型。在这一阶段中，我们需要决定如何将理论转换成方程。这就要求我们识别方程并且用足够的限制条件来使方程存在唯一一组参数来进行估计和解释。与此同时，我们还需要包括数据、方程以及研究者先验的理论信仰等方面的信息来决定是否正式接受这一选定的模型。[5]

建立理论和模型的第三个阶段是将选定的模型和具体的数据进行拟合并解释估计结果。我们发现，在这一过程中不存在太多的争论，因为我们在如何拟合模型以及用什么标准来判断模型的拟合程度等方面已经达成了广泛的共识（例如无偏性、有效性、最小均方误以及一致性等等）。但是一个相关的问题却需要我们引起注意，即从模型的估计中确定选定模型的动态特性。因为我们关注的是时间序列模型，所以我们需要关注这一方面的方法。

最后，在模型拟合以及解释结果之后，我们需要重返前几个步骤来评估我们在选定模型时所做的一些决定，并审视理论的哪些部分被支持，哪些部分没有得到支持。

模型建立过程中最关键的步骤是选定具体的函数形式（第二个步骤），其他

步骤在很大程度上都依赖于在这一过程中所作出的决定。在选定模型中确定一些变量或者确定错误的动态特征将会出现和最小二乘模型中一样的问题,即估计的有偏和无效。此外,如果我们没有包含多元系统中的已有的变量关系或重要因素,就会导致同时性偏误。也就是说,这些参数是有误的并且会在解释结果和假设检验过程中产生问题。

针对上述问题,最常用的解决方法是运用标准同时方程模型和单一时间序列模型。尽管这些模型能够解决很多问题,但是依然有局限性。我们接下来将讨论研究者在面对不同模型时应该如何作出权衡和选择。我们将介绍 4 种用于单一和多元时间序列数据的典型方法:自回归整合移动平均法(ARIMA)、同时方程或结构方程系统(SEQ)、误差纠正模型(ECMs)以及向量自回归(VAR)。在本书余下的部分中,我们交换使用同时方程和结构方程。我们将会讨论如何运用上述的每种方式来将动态同时关系模型化。当这种动态同时性关系能够或不能被理论、经验表述或者统计模型准确地设定时,我们将会帮助研究者作出特定的选择。

接下来的内容中最关键部分是,我们更多地强调方法而不是技术。上述所有方法都会运用各种不同的线性回归(最小二乘法、广义最小方差、多层最小方差等)或者最大似然估计。这些方法之间主要的差别是在进行统计推论和结果解释时的不同假设和缺陷。

同时方程方法

第一个建立时间序列多元方程模型的方法是同时方程。同时方程模型目前在社会科学的各个学科中广为运用。这种研究范式主要是由耶鲁大学在 20 世纪四五十年代发展起来的。这一委员会的初衷是发展出一套运用计量经济学来将经济问题模型化的方法论范式,所以研究者的工作主要是将当时已有的计量经济学方法用于经济学研究中的大规模、多层次方程模型。所以,早期的 Cowles 模型主要是凯恩斯(Keynesian)宏观经济学理论的经验表述。

建立同时方程的模型是基于将单个理论和方法进行经验表述,然后将其运

用于一系列方程。用单一理论来设定多个变量之间的关系，就必须确定哪些变量是外生于方程系统的，而哪些又是内生的。外生变量是那些由系统外因素决定或者被认为是固定不变的因素（在过去或当下的某个时点固定，但并不一定是恒定的），而另外一些由方程系统决定的变量和方程的因变量就是内生的。所以，同时方程模型的结果就是运用一个单一的结构方程系统来表述变量之间的关系。之所以只关注一个单个理论，是因为多个理论会导致不同的、非嵌套关系的结构方程设定（有关这一点，详见 Zellner, 1971; Zellner & Palm, 2004）。

现在，让我们回到政党参与率和政府政策支持率关系的例子中。用同时方程模型来表述这些变量之间的关系就需要两个方程，每个变量用一个方程。每个内生变量都是其他每个变量及其过去值的函数。对这个系统进行估计，首先需要做的是将方程系统重写为一组简化形式的方程，其中每一个内生变量都是一组预先确定的或者外生变量的函数。这种建模方式中未设定的因素是，如何决定有多少变量的过去值影响当下这个方程系统。或者说，如何将这个方程系统进行识别。典型的方式是由“理论”和假设检验的结果来指导我们包含或舍去某些变量。

用这种方式来建立同时方程模型会出现几个问题。首先，替代理论必须被嵌套在一个可比较的结构内。如果模型不能将其嵌套（由于非线性或者不同设定的问题），那么这个单一的系统将无法被用来比较不同模型。第二，模型要求在包含或舍去不同变量以及滞后变量时作出选择。两种常见的方法是，将“预先确定”或滞后内生变量限制为外生的，并且将变量只分为内生与外生两种。这里，“理论”指导我们限定模型的参数。我们通常的逻辑是用假设检验的结果来决定变量的取舍，但是这种方法会导致最终模型存在预检验偏误，因为基于检验结果而删除变量会使模型的拟合度非常好，从而让我们过于自信地估计结果。

正如西姆斯所指出的，这种舍弃变量的方法经常不被理论和经验分析所支持（Sims, 1980）。这将会导致以下结果，即额外的滞后变量被同时方程模型包含或者舍弃会引发错误的动态关系设定。即使该模型具有白噪音性质或非序列相关的参差，模型的设定也可能是错误的，并且暗示了错误的动态关系设定，

因为这种方法对参数空间进行了错误的限定。

最后，这些模型在预测政策效果上往往很差。相反，在这种情况下，一些简单的模型一般会优于复杂的、多方程的同时方程模型。

自回归整合移动平均模型

处理多元时间序列模型的另一个方法来自时间序列视角，这种方法将多元时间序列视为多个单时间序列的集合。在这个假设下，研究者可以运用标准的"博克斯—詹金斯法"或者自回归整合移动平均模型来处理每个单时间序列(Box & Jenkins, 1970)。当我们知道了变量之间的动态关系后，我们就可以建立模型并将一些变量视为冲击、扰动或者其他对自回归模型中自变量产生影响的外生效应。

博克斯—詹金斯方法旨在预测和描述一个时间序列的行为(Granger & Newbold, 1986)。基本的博克斯—詹金斯方法是定义一组模型——也就是自回归整合移动平均模型——来描述一个时间序列。接下来要做的就是将一系列自回归整合移动平均模型和每个时间序列拟合，这样做的目的是选择具有非相关残差的最简化模型。这一方法需要我们指定模型中的内生和外生变量。这一方法在进行预测时非常成功，实际上，博克斯—詹金斯方式的模型在进行预测时优于同时方程模型，主要原因是其简洁性，即模型的建立是运用简化原则并且尽量让数据说话。

运用博克斯—詹金斯方法来研究各政党参与率和政府政策支持率随时间变化的关系问题时，研究策略将按照如下进行：假设我们最感兴趣的是预测公众对某一政策的支持率。我们首先需要建立一个有关公众支持率动态变化的一元自回归整合移动平均模型。接下来，在模型确定后，我们需要做的是加入政党参与率这一协变量来看其是否提高了公众支持率模型的拟合度。方法之一是运用各种政党参与变量的测量方法(包括当下值和各种滞后值)，假设检验将会决定哪种测量是最佳的。另一种方法则是让模型去拟合特定的政党参与率的测量。

这种方法在建立多元时间序列模型时具有几个缺陷。首先，它忽视了这样一个事实，即模型中的一些变量可以代表另一些变量的动态变化。如果这种情况出现，那么估计过程将会导致严重的数据过度拟合，因为标准的博克斯—詹金斯方法是运用变量自身的过去值来进行解释。第二，这一方法首先关注的是变量的动态变化本身，而非变量在系统中的普遍关系。第三，由于估计是在不同的方程中进行的——也就是说，一个方程针对一个变量进行估计——因此我们可以认为，除非方程之间是绝对独立的，否则将会存在无效的估计。最后，除非变量是通过某种特定的方式因果相关的，否则将变量放在不同的自回归整合移动平均模型中将会导致无效的估计。原因是，如果一些变量的残差是同时相关的(在每一个相同的时点)，那么估计将是无效的。只有当每一个方程的估计结果都是明确的，我们才能运用一系列独立的方程来建立模型。

误差纠正模型和伦敦经济学院方法

误差纠正模型是自回归整合移动平均模型以及同时方程模型的一个特例。由于这种模型最初受到伦敦经济学院一些经济学家的推崇和发展，因此也被命名为“伦敦经济学院方法”(Pagan，1987)。建立误差纠正模型的方法是通过规定变量之间长期关系的方式来设定两个或多个变量的自回归分布滞后模型。

误差纠正模型和自回归整合移动平均模型的区别在于对变量之间长期关系的直接模型化，通常来说，这种关系包括随机趋势和决定趋势。在自回归整合移动平均模型中，这些长期因素、趋势或者单位根被区别对待，从而能建立一个稳定的数据序列来符合自回归整合移动平均模型的要求。但误差纠正模型则是将两个或多个数据序列中的长期因素作为彼此的函数。通过将两个或多个数据序列中的长期因素模型化，误差纠正模型就能获得这些序列的共有特征。利用这一共同特征，误差纠正模型会对所有变量生成一个共同的长期效应模型，并且辅以一个短期效应误差纠正机制，用来描述各个变量如何随着长期因素变化或平衡。

误差纠正模型可以被用于静态和非静态数据。对于静态数据，误差纠正模

型可以估计变量之间的普通或平衡关系，并且指出变量是如何围绕这个平衡而发生改变的。这一模型和自回归分布滞后模型的效果相同，后者其实是一个带有外生变量的自回归整合移动平均模型。对于非静态或趋势数据，误差纠正模型的建立需要从一组专门的数据序列开始——两个或多个含有单位根或一阶整合的数据序列。[6]我们可以用单位根检验的方法来确定这一组数据序列（例如增广迪基—福勒检验）。当确定一组数据序列为单位根后，就应该运用专门的估计方法来同时估计数据中变量的长期和短期关系。对于二元关系，可以运用一步或两步误差纠正模型。对于多元时间序列（尤其是具有单位根的数据序列），则需要引入误差纠正向量（VCEM）（详见 Johansen，1995）。这一过程的第一步是确定数据的普通随机趋势。随后，在对长期趋势进行估计之后，运用一个回归模型来估计长期趋势中的各个短期关系。

无论是误差纠正模型，还是多变量情况下的向量误差纠正模型，都是基于对一个多元时间序列回归模型中的长期和短期因素进行描述。研究者可以检验长期和短期动态过程中变量的各种关系以及这些关系是如何随时间发生变化的。对于非静态数据，误差纠正模型保证了时间序列中有一个特殊的“因果”关系（Engle & Granger，1987），这一因果关系被称为“格兰杰因果关系”，其中一个时间序列过去的值必须能够（线性地）预测其他序列的当下值。换句话说，在两个相互关联的时间序列中，变化趋势是由一个变量支配或预测的。这样一来，这些模型就成为同时方程模型的一个特殊例子，因为误差纠正模型是在不同的时间序列中假设并估计一个共同的时间趋势结构。

让我们再次考虑政党参与率和公众对政府政策的支持程度两个变量随时间变化的例子。有些观点认为，这些变量是单位根或非静态的，因为它们是多个序列中持续存在的事件和冲击的加总。如果是这种情况，那么误差纠正模型就是适用。误差纠正模型可以用于评价变量之间的长期和短期关系以及两个相互关联的变量中的格兰杰因果关系。误差纠正模型可以让研究者运用假设检验来决定多个时间序列中的长期和短期关系结构。利用误差纠正模型来表达变量之间的关系能够给我们提供更多的动态信息，但是我们必须估计各个短期效应的相互关联关系。

利用各种误差纠正模型的思想来建立统计模型并进行推论的方式虽然十分成熟(例如 Banerjee, Dolado, Glbraith & Hendry, 1993),但是对于非静态数据和单位根数据来说,将会极为复杂。许多经济学变量(消费量、国民生产总值、政府开支)都含有单位根。这是我们在思考这些模型时的一个重要原因,因为误差纠正模型能够让我们正确地看待变量之间长期和短期的动态关系,但是在非静态数据中利用这些模型进行频数推论时却很复杂,这是因为模型中的单位根变量会导致非标准分布以及动态分析计算的复杂性。这就意味着,在对误差纠正关系和误差纠正关系的数量进行假设检验,以及对含有非标准分布参数的模型进行检验时,必须用非标准检验统计表进行模拟和分析(Cromwell et al., 1994; Lutkepohl, 2004)。况且,误差纠正模型中的因果结构可能并不容易确定。西姆斯、斯托克和沃森指出,利用含有多个单位根的误差纠正模型进行推论是十分复杂的(Sims, Stock & Watson, 1990)。

向量自回归

对多元时间序列建立模型的最后一个方法是向量自回归模型。该模型的使用者并不假设自己已经知道多元时间序列背后的正确结构以及变量之间的潜在关系,取而代之的是关注时间序列中的潜在相关关系以及动态结构。

向量自回归方法在建立模型之初就关注时间序列之间相互关联的动态关系,并且思考以下一系列问题(与同时方程模型相反):

第一,我们怎么能够认为一些滞后变量不应被包含在每个方程内?或者说,对变量之间动态关系的识别加以限定的做法是否行得通?

第二,一个变量是如何通过时间因素来影响另外一个变量的?

第三,如果一个变量影响方程系统中的一个方程,我们怎么能够认为它不影响另一个或一些方程?

第四,一个合理的观点认为,一个变量可以被其刚刚过去的值加上一个随机项来最好地预测。在这个例子中,变量的过去值具有很小的预测价值,政策制定者和分析者感兴趣的往往是那些随机项,例如创新和政策冲击会引发什么

样的结果。在这个框架中,这些冲击是外生变量。

上述几条都是对同时方程模型标准方法的批评。向量自回归和同时方程模型的最主要的差别在于,它对方程系统中的所有时间序列做了一个完整的动态设定。向量自回归的模型化主要基于沃德分解定理(Hamilton, 1994:108—109; Wold, 1954)。沃德证明,每一个动态时间序列都可以被分割成一组固定项和随机项。

所有这些批评都指向对动态关系的理解。西姆斯对这些批评做出了回应(1972、1980),并基于将系统中的变量进行动态分解的思想,开创了向量自回归方法。他提出了3个拒绝使用标准同时方程模型的原因:

第一,利用同时方程模型时,在参数识别阶段设定限制并不是基于理论的,因此必然导致对模型结构以及估计的错误结论。

第二,同时方程模型一般基于对变量内生和外生性的弱化假设上。因为变量真正的滞后长度并不是预先已知的,所以随后的识别过程所建立的对内生性的假设可能是似是而非的。对动态同时方程模型的正式识别要求对所有变量滞后长度的准确把握,否则,有关识别的一些假设将无法被保持(Hatanaka, 1975)。

第三,如果模型中的变量本身就是政策引发的效应,方程识别就可能因为现实中的一些约束条件而产生新的问题。一个合理的批评是,虽然模型被假设在其他条件不变的情况下为真,但实际上,如果这些条件并不是不变的,那么我们就需要评价不同变量识别方式的概率意义。

对于同时方程模型的识别问题,西姆斯提出的处理方法是集中对模型简化形式采取动态识别。这种方法与同时方程模型方法的不同在于,后者关注的是确定模型过程中的识别选择,而西姆斯则是确保对多元时间序列数据建立模型时,能够为一些动态的序列提供完整的表达形式。我们可以运用多元自回归模型来解释所有的动态变量。[7]

西姆斯提出的向量自回归模型是一个通过多元自回归模型,将每个自变量回归到其自身及其他变量的过去值的方程系统。所以,向量自回归模型的建立就取决于对合适变量的选择(基于理论)。在处理方程动态结构的识别问题时,

我们必须运用样本数据来检验一个合适的滞后长度。西姆斯认为，向量自回归模型的一个批判性贡献在于，它可以为有关多元时间序列数据的经验争论定义一个合适的“战场”。因为该模型为一系列相关时间序列的动态性以及经验规律提供了一个具体的模型形式。从这一点上来说，我们可以以此为起点来发展及完善经验模型，进而评价理论的争论。

向量自回归模型的逻辑也可被用于政党参与率和公众对政府政策支持度的例子。建立向量自回归模型之前，有关简化形式的动态性的预设被置于中心地位。这与前三种方法有着明显的区别，因为同时方程模型是预设一组变量的结构关系，自回归整合移动平均模型则是预设一个动态过程，误差纠正模型是预设变量之间存在先验的因果关系。因此，向量自回归模型并没有对两个序列之间的关系强加一个可能的结构和动态关系，而是为两个序列各建立一个方程。每个变量都会被回归于其自身以及其他变量的过去值，剩余残差项则被认为是由外生冲击和创新导致的(在序列相关检验之后)。我们可以通过观察每个方程对这些外生冲击的反应来看这些因素如何影响观察到的方程系统。在考虑了这些动态过程之后，接下来要做的就是有关两个变量之间格兰杰因果关系的推断以及确定两个动态序列之间的内生结构。

利用向量自回归方法建立多元时间序列模型并不仅仅依靠单一的理论，而是在没有任何识别假设的前提下，利用多种理论的比较和评价(运用假设检验)的方法来确定模型。因为如果对方程的识别作出预设，那么就会和同时方程模型一样存在问题。由于向量自回归模型中的变量没有预先被分为内生变量和外生变量，我们就不会违反模型设定的错误，也不会将原本是内生的变量错误地设定为外生变量，进而引发同时性偏差的问题。

向量自回归模型和同时方程模型之间最关键的差别在于用不同的方式对待方程识别的预设。在同时方程模型中，方程的识别是固定不变的，是由单一的理论确定的。而在向量自回归方法中，诸如零阶限定这样的预设则被认为是不正确的(从一些方程中排除一些变量，或者在一些方程中省略一些变量的过去值)。因此，为了消除这些不正确的限定所产生的偏差，向量自回归模型被认为可以排除这些变量可能影响估计有效性的偏误。同时方程模型估计中的偏

误来源于排除了一些本应该包含在方程中的变量的过去值。按照西姆斯的话来说就是，在对同时方程模型进行识别的过程中，一些变量的部分过去值经常被错误地排除了。这些不正确的限定会导致忽略相关的滞后变量，进而产生忽略变量偏误。缓解这一问题的方法就是尽可能包含所有滞后变量和值（比必要情况下的更多）。这种做法是以牺牲估计有效性的方法来减少估计的偏误。

向量自回归模型的最主要识别预设是变量之间的同期效应如何互相影响。因为向量自回归模型是依据系统中变量的过去值来设定的，因此方程的识别主要关注对残差或者残差的同期协方差矩阵的设定。这样做的好处是，我们可以把对动态模型的解释和模型的识别区分开来。这一方法还可以让研究者清楚地看到，方程的识别与变量动态变化路径是如何相关联的。

向量自回归模型对数据与模型的相互作用有着不同的理解。向量自回归模型的目标是为数据的动态和相关关系提供一个概率模型（Sims，1980）。因此，在利用简单无偏的设定对数据动态关系以及模型的不确定性进行说明后，向量自回归模型的估计效果是最佳的。为了达到这个目的，预检验偏误必须被避免（Pagan，1987）。因此，与经典研究思路的“设定—估计—检验—再设定”逻辑不同，同时方程模型、自回归整合移动平均模型以及向量自回归模型很少用假设检验的方式来证明设定的正确性，这样会得到一个偏差较小的模型及其动态关系的表达式，而不会像其他建立模型的方法那样，虚假地认为得到了一个精确的模型设定。也就是说，一旦我们进入设定—检验的循环，那么得到的推论结果就是检验过程的一个函数，这个结果的可信度比直接进行模型估计后所报告的统计检验以及显著性水平或者 P 值要低。

比较和总结

以上对建立多元时间序列模型的各种可能方法的简单介绍旨在将各种方法与向量自回归方法衔接在一起（Pagan，1987；Sims，1996）。自回归整合移动平均模型、误差纠正模型和同时方程模型都是广义向量自回归模型的特例。弗里曼、威廉姆斯和林对向量自回归模型和同时方程模型进行了基本的对比

(Freeman, Willianms & Lin)。而本书对不同的模型进行讨论的目的是比较动态关系如何被模型化以及如何进行推论。

表 2.1 对每一个方法进行了概括，对弗里曼、威廉姆斯和林最初的概括进行了拓展。表格呈现了不同的方法在设定时间序列模型时的方法差异。

最关键的一点是，向量自回归是对其他方法的归纳。其他 3 种模型建立的方法只是关注时间序列数据的某些特征，这些特征在实际操作中可能是正确的。但从模型建立和理论检验的观点出发，向量自回归模型则更具有普遍性。

表 2.1　时间序列各种建模方法的比较

	自回归移动平均整合	误差纠正	结构方程	向量自回归
建模设定	单一理论，着眼于一元时间序列	对协同整合关系和单位根进行检验后，设定长期和短期趋势的动态关系	单一理论以及对内生性和外生性的假定	建立在多个理论基础上并包含多个内生变量
估计	最大似然估计、最小二乘估计	乔纳森过程分析、一阶段或双阶段回归	高阶最小二乘估计和最大似然估计、异方差纠正和序列相关、对正交和过度识别的检验	最小二乘估计和对滞后时长的检验
方法论传统假设检验	对系数进行单独分析	对协同整合关系进行检验并对短期动态关系进行检验	对系数进行单独分析，检验模型的拟合优度	对一组系数进行显著性检验，对外生性进行检验
动态分析	动态乘子和干预分析	协同整合向量分析和冲击反应分析	模拟和模型动态的推断	预测、模型推测、预测误差方差分解和冲击反应分析

那么，我们为什么推崇向量自回归模型呢？首先，我们并不想排除结构方程模型。实际上，当对模型作出的限定条件与数据和现实相吻合的时候，那么结构方程模型将是一种很好的估计方法。它能够帮助我们更好地进行推论，较好地概括数据的动态特征以及变量关系的表述。其次，结构方程的建模过程是

明确遵照动态关系中变量之间“刺激—反应”关系的假设，因此，按照此假设进行阶段性的建模能够生成一个较为精简的模型。

那么当结构方程模型出现问题时，我们是否应该用向量自回归模型来替代它呢？在以下 3 种情况下，我们应该这样做：首先，我们已经知道或者检验结果表明，模型中的变量关系结构不符合结构方程模型的设定。其次，当我们要对政策的反事实进行分析时，除非结构方程模型的设定是正确的，否则我们很容易得到错误的推论结果。最后，如果我们的目标是分析一些不确定的动态关系，那么向量自回归模型肯定优于结构方程，因为它不太可能利用一些特设的预检验而对未知的动态关系作出过于精确的假设。

最后要说的是，在一些情况下，我们可能更倾向于用误差纠正模型。例如，我们想同时分离一些时间序列的长期和短期行为，或者当一些趋势变量或根变量存在于一个多元时间序列模型中。在这些情况中，我们实际上还是在运用向量自回归模型，只不过是对多元时间序列模型中含有的长期行为提出一组限定条件或者预设。本书的目的是让研究者懂得如何运用误差纠正模型和向量误差纠正模型。后者是一个更具普遍意义的非限定性向量自回归模型。我们将在下一章继续讨论向量自回归模型和误差纠正模型的关系。

在下一章中，我们概括了向量自回归模型的一些数学特征。接下来讨论如何运用该模型对多元时间序列数据的关系进行推论。

第 3 章 | 基本的向量自回归模型

向量自回归模型并非一个统计技术或者方法论，而是对一系列（内生）变量进行动态建模的方法。这一方法是运用多元回归以及多元相依回归来关注多元时间序列的动态性。我们最关心的是数据及其动态性。向量自回归模型的中心要义是，我们必须抱着怀疑的态度来看待数据之间或模型参数之间的关系。

如何持有怀疑的态度？让我们假设一个充满动态关系和相关结构的时间序列数据。请思考我们是如何"看"这个数据的。如果我们用双眼去看，即拥有完整的信息时，那么就可以预先知道数据中蕴含的所有动态关系。但是当我们闭上一只眼睛，即信息相对缺乏时，那么就只能看到部分动态信息，从而失去对数据的整体把握（能看到何种动态关系取决于你看到了哪部分已知信息）。向量自回归模型努力地让我们睁开双眼，不会错误地闭上眼睛，或者被不正确的预设阻碍视线。

那么，向量自回归模型到底是什么呢？简单地说，就是一个相互依存的简化动态模型。对于方程系统中的每一个内生变量都建立一个方程，使这些变量成为其自身过去值以及其他内生变量过去值的函数。一般来说，在每个方程中，每个变量的滞后或者过去值数量是相等的。其他外生变量或者控制变量则会被包含进方程作为额外的自变量。

在本章中，我们将呈现设定和解释一个基本的向量自回归模型时所需的数学细节。我们假设读者仅具有对线性回归模型和代数矩阵的基本知识。从这个基础出发，我们将讨论向量自回归模型的一些特性。

本章有两个目的。第一个目的是展示向量自回归模型如何与其他大家熟

知的模型相联系，例如结构方程模型。第二，我们将列出设定、估计和解释向量自回归模型时所需的基本术语和技巧。在第 4 章中，我们将会基于这些讨论举两个实例。

本章的内容是按照如下顺序进行组织的：首先，我们列出一个动态同时方程模型的通式和一个向量自回归模型，通过比较，我们将呈现后者如何帮助我们“不去损害”我们对数据的理解。接下来，我们将讨论向量自回归模型中主要的模型设定和推论决策。随后，我们将讨论一些细节问题，包括对滞后长度的选择、估计以及向量自回归模型的统计推断。另外，本章对动态反应分析给予了特殊的关注，我们将通过冲击相应矩阵以及预测误差方差分解等方式来达到这一目的。最后，我们还讨论了设定向量自回归模型过程中几个极易被忽视但却被认为是标准做法的步骤。

动态结构方程模型

第 2 章概述了动态结构方程模型在计量经济学和社会科学领域的发展历程和基本贡献。在本章中，我们将专门讨论这一模型。我们设想基本动态结构方程模型包含两个内生变量 Y_t 和 Z_t。每个变量都在时间 1，…，T 被观测到。两个变量的滞后值或过去值分别记为 Y_{t-l} 和 Z_{t-l}，其中 $l=1, 2,\cdots$，代表在 t 时间之前 l 个阶段所观察到的值。

那么，含有这两个变量的动态结构方程模型系统表达式如下：

$$Y_t = \alpha Z_t + \gamma_{11} Y_{t-1} + \gamma_{12} Y_{t-1} + u_{1t} \qquad [3.1]$$

$$Z_t = \theta Y_t + \gamma_{21} Y_{t-1} + \gamma_{22} Z_{t-1} + u_{2t} \qquad [3.2]$$

其中，

$$u_{it} \sim N\left(0, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right)$$

之所以称之为一个同时方程系统，是因为模型中的所有方程都决定了至少一个内生变量的值。换句话说，我们可以在每个方程中看到每个变量的同期值——

Z_t 在方程 3.1 的右边，Y_t 在方程 3.2 的右边。模型的同时性来源于这样一个事实，即每个变量都依赖模型中其他变量的同期值，而动态关系则取决于滞后值。

我们将 Y_t 和 Z_t 视为内生变量。那些在 t 时段的模型中就已经被确定的变量（例如 Y_{t-1} 和 Z_{t-1}）既可以被视为外生变量，也可以被视为取值已知的滞后内生变量。请注意，这个模型既概括了与内生变量相关的变量的概括，又包含了这些变量之间的时间关系。

方程 3.1 和方程 3.2 中的模型可以被看做结构方程形式的模型。需要注意的是，要对这个系统中的方程进行估计，我们必须用一个方程替换另一个，因为我们至少需要一个方程才能确定另一个方程。并且，我们不能将两个方程分开进行最小二乘法估计，因为忽略一个方程代表忽略了变量之间的同时性，这会导致回归估计中的同时性偏差。而且两个变量的系统是动态的，即每个变量在 t 时段的取值取决于变量在 $t-1$ 时段的值。这就是对变量进行一阶自回归的过程。

为了产生一组可被一致估计的方程，我们采取简化形式。我们将方程3.2带入方程 3.1 来求解 Y_t，这就得到了 Y_t 的下列方程：

$$\begin{aligned} Y_t &= \alpha[\theta Y_t + \gamma_{21}Y_{t-1} + \gamma_{22}Z_{t-1} + u_{2t}] + \gamma_{11}Y_{t-1} + \gamma_{12}Z_{t-1} + u_{1t} \\ &= \alpha\theta Y_t + \alpha\gamma_{21}Y_{t-1} + \alpha\gamma_{22}Z_{t-1} + \alpha u_{2t} + \gamma_{11}Y_{t-1} + \gamma_{12}Z_{t-1} + u_{1t} \end{aligned}$$

$$Y_t(1-\alpha\theta) = (\alpha\gamma_{21} + \gamma_{11})Y_{t-1} + (\alpha\gamma_{22} + \gamma_{12})Z_{t-1} + \alpha u_{2t} + u_{1t}$$

$$Y_t = \frac{(\alpha\gamma_{21} + \gamma_{11})}{(1-\alpha\theta)}Y_{t-1} + \frac{(\alpha\gamma_{22} + \gamma_{12})}{(1-\alpha\theta)}Z_{t-1} + \frac{(\alpha u_{2t} + u_{1t})}{(1-\alpha\theta)}$$

对 Z_t 的简化方程也可以得到：

$$\begin{aligned} Z_t &= \theta[\alpha Z_t + \gamma_{11}Y_{t-1} + \gamma_{12}Z_{t-1} + u_{1t}] + \gamma_{21}Y_{t-1} + \gamma_{22}Z_{t-1} + u_{2t} \\ &= \theta\alpha Z_t + \theta\gamma_{11}Y_{t-1} + \theta\gamma_{12}Z_{t-1} + \theta u_{1t} + \gamma_{21}Y_{t-1} + \gamma_{22}Z_{t-1} + u_{2t} \end{aligned}$$

$$Z_t(1-\theta\alpha) = (\theta\gamma_{11} + \gamma_{21})Y_{t-1} + (\theta\gamma_{12} + \gamma_{22})Z_{t-1} + \theta u_{1t} + u_{2t}$$

$$Z_t = \frac{(\theta\gamma_{11} + \gamma_{21})}{(1-\theta\alpha)}Y_{t-1} + \frac{(\theta\gamma_{12} + \gamma_{22})}{(1-\theta\alpha)}Z_{t-1} + \frac{(\theta u_{1t} + u_{2t})}{(1-\theta\alpha)}$$

这一求解后的系统就是我们所知道的模型的简化形式。这些方程清楚地展现了内生变量如何与预先确定的变量相关联。简化形式可以被写成下列更为紧凑的方式：

$$Y_t = \Pi_{11} Y_{t-1} + \Pi_{12} Z_{t-1} + \varepsilon_{1t} \qquad [3.3]$$

$$Y_t = \Pi_{21} Y_{t-1} + \Pi_{22} Z_{t-1} + \varepsilon_{2t} \qquad [3.4]$$

其中，Π_{ij} 代表在前一个简化方程中，第 i 个方程中变量 j 的系数。这个对 Y_t 和 Z_t 的简化系统可以被最小二乘法估计。但是用最小二乘法估计得到的简化形式系数 Π_{ij} 并不是我们依据理论设定的结构方程 3.1 和方程 3.1 的参数。结构方程的参数必须通过简化形式来进一步获得。我们有 6 个结构方程参数，简化模型却只能估计出 4 个参数，但是从 4 个参数的模型恢复到 6 个参数的模型是十分困难的。这时研究者面临这样一个选择，因为在这种情况下，我们需要作出一些预设来识别模型。这些预设将会影响我们有关结构方程参数的推断以及对数据动态性的描述。

为了更好地说明这一问题，我们假设研究者需要从简化模型的参数来对方程 3.1 中的参数进行恢复，从而得到一个对 α 的一致估计。对 α 的一致估计有一个前提条件，即方程 3.2 不能对方程 3.1 产生与其参数相关的影响。用数学形式来表达，就是 $\theta = 0$，所以 $E(Z_t u_{1t})$ 也将为 0。并且，为了知道对参数 Π_{21} 的最小二乘估计是否满足一致性条件，我们还需要知道 Π_{21} 是否等于 0。换句话说，就是方程系统没有影响因素通过变量的过去值而使对 α 的最小二乘估计无效。但是，知道简化模型参数 Π_{21}，并不能说明模型是否满足这一前提条件 $(\theta = 0)$。我们需要这一条件来使得对 α 的估计满足一致性原则。

让我们来思考能够决定 Π_{21} 和 θ 关系的 3 个识别预设：(1) $\gamma_{21} \neq 0$，但是 $\theta = 0$；(2) $\gamma_{21} = -\theta\gamma_{11}$，所以 $\Pi_{21} = 0$，但是 $\theta \neq 0$；(3) $\gamma_{21} = 0$ 并且 $\gamma_{11} = 0$，但是 $\theta \neq 0$。在第一种情况下，我们假设只有 Y_t 的过去值可被用来预测 Z_t。所以这将不存在方程间的相互影响，但是 Π_{21} 的值将会是 γ_{21}。在第二种情况下，我们假设 Y_t 的过去值和现在值的系数不成比例地相互抵消，因此我们无法用其预测 Y_t 和 Z_t。但是在这种情况下，$\Pi_{21} = 0$ 且 $\theta \neq 0$，所以对 α 的估计将不满足一致性

原则。在最后一种情况下，我们认为，虽然 θ 不为 0，但是产生的简化参数却是 0。

上述讨论告诉我们，如何确定简化模型和结构模型之间的参数关系对于解释估计结果的有效性起着至关重要的作用。如何从简化参数恢复结构方程参数的识别决定也对同时方程系统的动态性产生约束。例如在上述第一种情况下，我们假设（有可能是错误的）只有 Y_t 的过去值可以预测 Z_t。在第二种情况下，两个变量之间的同期相关使得 Y_t 的过去值会在两个方程中相互抵消，因此对所描述的模型动态关系产生限制。在最后一种情况下，Y_t 在两个方程中都没有预测值，Y_t 对 Z_t 的所有解释力都仅仅来自同期值。所有这些识别限定都为方程系统的动态性带来了限制。

面对这些表述 Y_t 和 Z_t 关系的极为不同的动态模型，我们就必须借助一些理论来识别方程，而非仅仅追求估计的一致性。但我们其实可以避免这些选择。想想我们在第 2 章中讨论过的替代解释。我们可以在不对结构方程模型进行有可能错误的限定时，只分析简化模型。在这种情况下，我们关注的是变量之间的动态关系，并且允许各种可能的同期关系的出现。

向量自回归的简化形式

为了替代这种存在错误识别假设的同时方程模型，西姆斯提出应该直接对简化模型进行分析(1972、1980)。因为我们对时间序列的分析主要是关注方程系统的动态性，因此直接分析简化模型并不存在问题。实际上，简化模型使我们在评估模型中感兴趣的变量的动态性之后，能够对各种识别预设作出检验。

在针对动态系统的向量自回归模型中，我们将内生变量的系统写成系统中其他既定变量及其已知值的函数。此时的向量自回归模型就是一个由未加限定的简化方程组成的系统。让我们写一个具有 m 个内生变量的方程系统，用 y_{it} 来表示第 i 个自变量，其中 $i = 1, 2, \cdots, m$，t 则是时间指标。如果将向量自回归模型的简化模式写成标量形式，那么它就应当包含如下方程：

$$
\begin{aligned}
y_{1t} &= \beta_{10} + \beta_{11} y_{1,t-1} + \cdots + \beta_{1,p} y_{1,t-p} + \cdots \\
&\quad + \beta_{1,m} y_{m,t-1} + \cdots + \beta_{1,mp} y_{m,t-p} + e_{1t} \\
y_{2t} &= \beta_{20} + \beta_{21} y_{1,t-1} + \cdots + \beta_{2,p} y_{1,t-p} + \cdots \\
&\quad + \beta_{2,m} y_{m,t-1} + \cdots + \beta_{2,mp} y_{m,t-p} + e_{2t} \\
&\vdots \\
y_{mt} &= \beta_{m0} + \beta_{m1} y_{1,t-1} + \cdots + \beta_{m,p} y_{1,t-p} + \cdots \\
&\quad + \beta_{m,m} y_{m,t-1} + \cdots + \beta_{m,mp} y_{m,t-p} + e_{mt}
\end{aligned}
\tag{3.5}
$$

对于每一个内生变量，我们都将其回归到自身及其他内生变量的 p 个滞后值。因此，每个方程都包含 $mp+1$ 个回归系数，而整个方程系统总共有 $m(mp+1)=m^2p+m$ 个回归系数。为了方程的完整性，我们还必须对残差的分布作出假设。在这种情况下，我们假设残差的联合分布是正态的，或者 $e_t \sim N(0, \Sigma \otimes I)$，其中 $e_t=(e_{1t}, \cdots, e_{mt})$，$\Sigma$ 是残差项的一个 $m \times m$ 协方差矩阵，I 是 $T \times T$ 的单位矩阵，$\otimes$则是两个矩阵克罗尼克(Kronecker)乘积的一个乘子。[8] 这个模型是一个多元回归模型，其中所有的内生变量都被放在方程的左边，而所有既定的滞后变量都被放在方程的右边。

我们经常可以看到对上述方程 3.5 的另一种表达形式，是用矩阵来表示方程系统。在下面这个表达式中，m 个 t 时点的内生变量是一个$(1 \times m)$的向量和 $y_t=(y_{1t}, y_{2t}, \cdots, y_{mt})$，系数 β 则用矩阵 B_l 表示：

$$
y_t = c + \sum_{l=1}^{p} y_{t-l} B_l + e_t \tag{3.6}
$$

表达式中的 c 是截距(β_{i0})的向量，y_{t-l}是 l 个滞后变量的 $1 \times m$ 向量，B_l的 $m \times m$ 矩阵是第 l 个滞后项的系数($\beta_{i,1}$到 $\beta_{m,mp}$)。最后，e_t 是 $1 \times m$ 向量的残差。

方程 3.5 的最后一种表达方式可以看做一个多元回归模型的特殊形式。该模型可以被写成矩阵形式：

$$
\boldsymbol{Y} = \begin{pmatrix} y_{11} & \cdots & y_{m1} \\ \vdots & \ddots & \vdots \\ y_{1T} & \cdots & y_{mT} \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} y_{1,t-1} & \cdots & y_{mp,t-p} \\ \vdots & \ddots & \vdots \\ y_{1,T-1} & \cdots & y_{mp,T-p} \end{pmatrix}
$$

其中用数学符号来表示从 1 到 T 的所有观察值，所以模型就可以写为：

$$\boldsymbol{Y} = \boldsymbol{XB} + e \qquad [3.7]$$

其中，$\boldsymbol{Y}$ 是内生变量从 1 到 T 时点的 $T\times m$ 矩阵，$\boldsymbol{X}$ 是滞后内生变量的 $T\times(mp+1)$ 矩阵，而 $\boldsymbol{B}$ 则是将所有自回归矩阵合并后的$(mp+1)\times m$ 矩阵，对于 B_l，$l=1, 2, \cdots, p$。

请注意，向量自回归模型的简化形式和从方程 3.3 中推导出来的同时方程模型的简化形式是一样的。在方程 3.3、方程 3.5、方程 3.6 中，内生变量在 t 时点的值是因变量，而解释变量则是所有的滞后项。因此，我们没有武断地限定方程到底应该包含多少个滞后项，因为每个方程所包含的滞后项的数量是相等的。在下一节中，我们将更为正式地展示如何将向量自回归模型的简化形式作为广义的动态同时方程模型。同时，我们还要讨论如何对这一简化形式进行估计以及推论。在随后的章节中，我们将方程 3.6 和方程 3.7 作为向量自回归模型的表达式。

向量自回归模型与动态同时方程模型的关系

在本节中，我们将讨论如何将向量自回归模型作为广义动态同时方程模型的表达形式。在这之前，我们要先定义一个动态同时方程模型更加普遍的形式。如果我们要用向量形式写出下列动态方程系统：

$$y_t A_0 = d + y_{t-1}A_1 + y_{t-2}A_2 + \cdots + y_{t-p}A_p + u_t \qquad [3.8]$$

其中，参数和数据的行向量被定义为：

$$y_t = (y_{1t}, y_{2t}, \cdots, y_{mt})$$
$$u_t = (u_{1t}, u_{2t}, \cdots, u_{mt})$$
$$d = (d_1, \cdots, d_m)$$
$$u_t \sim N(0, I)$$

其中，残差形式以一个均值为 0 的 $m\times m$ 单位协方差矩阵表示。A_i 矩阵是一个

$m \times m$ 矩阵，代表系统内内生变量的滞后值带来的影响。A_0 矩阵则代表内生变量之间的同期关系。为了确保系统的识别，我们要求 A_0 必须是一个可逆的满秩矩阵，即存在 A_0^{-1}。

如果我们在方程 3.8 的两端同时乘以 A_0^{-1}，那么可得到方程 3.9：

$$y_t = c + y_{t-1}B_1 + y_{t-2}B_2 + \cdots + y_{t-p}B_p + e_t \qquad [3.9]$$

其中，$c = dA_0^{-1}$，$B_i = A_iA_0^{-1}$，$i = 1, 2, \cdots, p$，$e_t = u_tA_0^{-1}$。方程 3.9 是方程3.8的简化形式。这与广义向量自回归模型的方程 3.6 具有相同的形式，即内生变量的同期值被回归到其自身以及系统内其他变量的过去值。

向量自回归模型简化形式中的同期关系实际上是对残差的协方差进行参数化的一个步骤，因为对同期关系的识别或正交是通过对 A_0 矩阵的参数化来完成的。为了展示结构方程中的同期关系如何成为残差简化形式的一个组成部分，我们先计算出残差项 e_t 的协方差矩阵，Σ 是矩阵 A_0 的同期结构函数：

$$\begin{aligned}\Sigma = V[e_t] = E[e_t'e_t] = E[A^{-1\prime}{}_0u_t'u_tA_0^{-1}] \\ = A_0^{-1\prime}E[u_t'u_t]A_0^{-1} = A_0^{-1\prime}IA_0^{-1} = A_0^{-1\prime}A_0^{-1}\end{aligned}$$

向量自回归模型的一个重要结论是对同期关系加以限定。换句话说，同时方程模型中的 A_0 矩阵是由向量自回归模型残差协方差的关系决定的。

我们为什么要将同期关系包含在残差项的方差中呢？在识别同时方程模型时，这种同期关系恰恰是排除一些变量的根据。但是正如我们在第 2 章中所说的，向量自回归模型是非限定性的简化形式，它包含所有有意义的变量，且不做任何可能错误的识别假设，而是对动态关系采取无限制的表述形式。我们可以用估计出来的模型来评价结构方程和对动态关系的设定，而不是用预先设定的参数来误导我们的推论。

模型的运用

我们如何运用向量自回归模型呢？由于该模型和特定的结构模型紧密联

系，那么其目的是什么呢？一般来说，研究者运用向量自回归模型主要是为了完成以下几件事：第一，确定内生变量之间的因果关系；第二，模型中一个变量的变化对其他变量的动态影响；第三，每个变量的方差有多少可以被自身变化所解释，又有多少可以被其他变量所解释？

第一件事可以运用格兰杰因果的思路来确定。为了确定格兰杰因果关系，必须运用一个假设检验来确定一个变量对预测另一个变量而言是否具有统计上的有效性。如果可以，我们接下来要做的就是基于数据关系来建构因果顺序。当然，必须对这些数据关系作出更具体的识别假设。这一检验的价值在于，它实现了将社会科学领域的许多假设归结为对变量之间外生性的探求。格兰杰因果思想通过将变量 X 放入对 Y 的预测方程中，从而将 X 的外生性与 Y 的预测值联系在一起。

第二，评价一个变量对其他变量的动态影响，可以看成是在一个单方程的时间序列分析中测量方程右边变量的变化所带来的长期、短期影响乘数。多方程的情况类似于冲击反应函数（IRF）或者移动平均反应分析（MAR）。这些多元动态乘数可以被用于检验变量之间是否存在动态的因果关系。通过转化向量自回归方程系统求解移动平均式的过程，我们可以得到其冲击反应函数（转化取决于识别假设）。我们之所以要对移动平均式进行转化，是因为这样可以看到外生冲击对方程系统的影响，并且找出方程是如何对这些冲击作出反应的。我们还可以通过移动平均式来分析对内生变量的同期相关关系所做的不同的识别假设。

评价向量自回归模型的最后一个方法是确定每个变量的方差有多少可以归因于其他变量的动态变化。这一过程也称为“创新计算”或者“预测误差方差分解”。运用这种方法时，我们必须知道变量的预测方差中有多少是因为变量 i，又有多少可以归于变量 j。这就使我们能够观察一个变量的变化如何导致另一个变量的变化。一个变量的预测误差越多地被另一个变量解释，那么后者对于预测和解释前者来说就具有越重要的意义。在这里，我们是想知道变量的预测动态性有多少来源于变量间的同期关系，又有多少来源于方程系统的动态性。

上述3种对向量自回归分析的可能性解释是为了让我们在简化模型的基础上更好地理解变量变化所产生的动态影响。与结构方程模型专注于具体的参数不同(这些参数有可能被错误地设定),向量自回归只是尽可能地描绘模型的动态性。上述所有模型都是在探讨这样一个问题:一组相关的时间序列变量如何在不同时间段上相互影响。

在介绍向量自回归模型的解释方法之前,我们先对其设定与估计的细节做一些讨论。随后我们将回到格兰杰因果、冲击反应分析以及创新计算等话题。

向量自回归模型的设定与分析

在对向量自回归模型系统中的方程进行设定的最初阶段,研究者会面临一些与标准的自回归整合平均移动模型或博克斯—詹金斯单时间序列分析不同的问题。首先,数据的多元性使得一些检验方法变得更为复杂甚至无效,例如设定搜索时的自相关函数。其次,对向量自回归模型简化形式的估计需要设定一个滞后时长 p。第三,解释外界冲击对方程的动态影响需要我们对所有变量的同期关系作出识别。最后,对系统的设定决定着向量自回归模型的估计方法。我们会在随后的章节中处理这些问题。

向量自回归模型的估计

对于非限定性向量自回归模型来说,方程3.7的最大似然估计是:

$$\boldsymbol{B} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad [3.10]$$

请注意,$\boldsymbol{B}$ 是一个 $(mp+1)\times m$ 的回归系数矩阵,其中,j 列代表 j 个变量的回归系数。由于残差的误差协方差被假定为分块矩阵,所以它看起来与回归模型是无关的。因此,我们并不需要马上估计 m 个方程,而是一个接一个地运用最小二乘估计来获得一致性的估计值。[9] 在这些情况下,由于没有对向量自回归的系数进行限定,估计是通过对 m 个方程进行逐步最小二乘估计来完成的,每个方程都有一个模型中的因变量。

残差的协方差矩阵可以被看成样本残差：

$$\hat{\Sigma} = \frac{1}{T}\sum_{t-1}^{T} \hat{e}_t{}'\hat{e}_t, \qquad [3.11]$$

其中，$\hat{e}_t$是方程 3.6 中多元回归后残差的 $1 \times m$ 矩阵。这是所有观测值的样本误差协方差矩阵。

滞后时长设定

标准的向量自回归模型系数和估计值取决于对滞后时长的设定。虽然在前面的讨论中有所涉及，但是我们必须为具有不同滞后时长的模型选定一个合适的统一滞后时长 p。

主要有两种方法可以被用于检验向量自回归模型的滞后时长。第一种是基于过往经验（特别是经济的周期循环）和数据的周期性。第二种则基于正式的假设检验。

向量自回归模型滞后时长设定的经验法则

利用经验法则，向量自回归模型已经包含了足够的滞后值来把握数据的完整周期。对于月度数据，至少包含 12 个滞后值。更典型的例子是我们还拥有跨年的季节性数据，这样就可以运用 13 到 15 个滞后值。对于季度数据，我们一般采用 6 个滞后值。这对于处理周期性的年度数据后余下的一些季度数据而言已经足够。为了确保估计的稳健性，我们可以利用下文将要讨论的检验方法，以最多 8 到 10 个滞后值来对模型进行评价。对于月度或者季度数据，这些滞后时长足够处理主要的季节性。这一点非常重要，因为即使是在非季节性数据中，也依然存在一些需要被模型化的季节趋势。[10]

最后一个经验法则是，选择的滞后时长不能大于任何一个方程自由度的 1/4。因此，如果一个时间序列拥有 120 个时点的数据，那么滞后值的数量就不能大于 $mp+1<T$，其中，m 是内生变量的数量，p 是滞后长度，T 是观察值的总数。所以，如果系统中有 3 个变量，那么滞后时长应该满足 $3p+1<120$ 或者 $p<119/3$，大约是 40 个滞后值。之所以作出这样的限制，是出于两个原因：其一，向量自回归模型的自由度太低会导致估计的相对无效；其次，在估计中运用过

多的滞后值会导致最小二乘法无法计算出系数。[11]

尽管这些规则非常概括，但它们是我们选定滞后时长的起点。对于非月度和季度数据，我们的方法并不适用。接下来，我们将运用假设检验的方法对 p 和 $p-1$ 的滞后时长的不同设定进行更为正规的评估。

向量自回归模型滞后时长设定的检验

两个经典的统计检验可以被用来评估向量自回归模型的滞后时长设定。第一个是基于似然比检验，即比较含有 p 个滞后值的模型与含有 $p-1$ 个滞后值的模型所达到的最大似然值。向量自回归模型的最大似然函数可以被写成：

$$L(\hat{\Sigma}, \boldsymbol{B}, p) = -\frac{Tm}{2}\log(2\pi) + \frac{T}{2}\log|\hat{\Sigma}^{-1}| - \frac{Tm}{2} \qquad [3.12]$$

其中，$\hat{\Sigma}^{-1}$ 是估计误差协方差矩阵的转置矩阵(见方程 3.11)。而 $\log|\hat{\Sigma}^{-1}|$ 是 $\hat{\Sigma}^{-1}$ 决定值的对数形式，即误差协方差矩阵的转置矩阵。似然函数中的 $\boldsymbol{B}$ 取决于滞后长度 p。因此，我们可以十分明确地利用 $L(\hat{\Sigma}, \boldsymbol{B}, \mathrm{p})$来定义最大似然函数。

最正式的检验方法是为两个含有不同滞后时长的模型构造各自的似然比检验值或卡方检验值，然后再进行比较。这里的卡方检验十分重要，因为它可以帮助我们确定向量自回归模型是否解释了数据中所有的动态关系，但是这一检验在渐进条件(T 趋向于正无穷)下才比较准确。模型检验的虚无假设以及备择假设如下：

H_0:虚无模型　向量自回归模型含有 $p = p_0$ 个滞后值

H_1:备择模型　向量自回归含有 $p = p_1$ 个滞后值 ($p_1 > p_0$)

对含有 $p_1 > p_0$ 个滞后值的模型进行似然比检验可以写成如下形式：

$$(T-1-mp_1)(\log|\hat{\Sigma}_0| - \log|\hat{\Sigma}_1|) \qquad [3.13]$$

其中，$\hat{\Sigma}_i$ 是含有 p_i滞后值的向量自回归模型的误差协方差。[12]这一检验值服从 $m^2(p_1 - p_0)$ 的卡方分布。自由度可以通过如下方法得出，即从每个方程的每个变量中除去 $p_1 - p_0$ 个滞后值所造成的影响。因为模型中共有 m 个内生变

量，所以假设检验中加以限定的模型减少了 $m^2(p_1-p_0)$ 个变量。之所以会发生这种情况，是因为假设检验中的模型被限定只能包含少于 $m^2(p_1-p_0)$ 个滞后值。

我们经常对这一检验进行小样本纠正。第一，检验是以 F 检验的形式给出的。实际上，这是在给定自由度的情况下，对卡方统计量进行重构，所以在渐进条件下，两者是相同的。第二，在计算最小二乘估计的标准差或者对样本方差的计算进行小样本纠正时，一般的做法是对向量自回归模型中每个方程的估计参数都进行似然比卡方检验。请注意，我们已经在上文中提及如何进行误差纠正。在运用方程 3.13 进行似然比检验时，用 T，也就是观测值的个数，减去非限定模型（mp_1+1）中每个方程的参数值。这样做是为了降低对模型中参数估计的卡方值的大小。这一调整后的统计量服从自由度为 $m^2(p_1-p_0)$ 的卡方分布。这种纠正方法是由西姆斯提出的。

对滞后时长进行假设检验经常会遇到一个很复杂的情况：加入方程的参数越多（例如滞后项），最终的似然比的值就越大，相应的拟合度也就越高。这与线性模型中加入更多的自变量会提高模型的拟合优度是一个道理。在这种情况下，似然比检验就可能会选择不正确的滞后时长，因为当似然数随着滞后项的增加而增加时，似然比就不能反映更多的滞后项所带来的影响。在这种情况下，模型的检验就需要对我们进行滞后时长的时间序列检验的预检验偏差进行进一步的纠正（Lutkepohl，1985、2005）。预检验偏差产生的原因是，我们所估计的是一个序列的检验，所以 p 对 $p+1$ 的检验依赖于 $p-1$ 对 p 检验的结果。这就导致我们应该运用一个更大的临界值或拒绝概率来进行高阶检验。[13] 换句话说，当我们需要在向量自回归模型中加入一个滞后项时，必须提出足够的理由和证据。

对滞后时长进行检验的第二个常用方法是利用信息准则，例如赤池信息准则（AIC）、贝叶斯信息准则（BIC 或 SC）以及汉南—昆信息准则（HQ）。信息准则是一种在模型拟合度和模型简化程度之间进行权衡的方法。上述信息准则都建立在一个模型似然函数的基础上，并将该模型包含的参数数量作为惩罚因子。对于两个拟合度相等的模型来说（即两个模型具有相同的似然值），惩罚因

子较小的更简化模型将会被信息准则认定为更优的选择。

AIC、BIC 和 HQ 信息准则在如何将参数确定为惩罚因子时存在着不同。计算 3 个信息准则的表达式如下，我们假设所计算的是滞后时长为 $p=0,\cdots,p_{\max}$ 的非限定向量自回归模型。

$$\mathrm{AIC}(p)=T\log|\hat{\Sigma}|+2(m^2p+m) \tag{3.14}$$

$$\mathrm{BIC}(p)=T\log|\hat{\Sigma}|+\log(T)(m^2p+m) \tag{3.15}$$

$$\mathrm{HQ}(p)=T\log|\hat{\Sigma}|+2\log(\log(T))(m^2p+m) \tag{3.16}$$

如果该模型所包含的样本容量为 T，且包含 $p_{\max}$ 个滞后项，那么 $\log|\hat{\Sigma}|$ 就是从方程 3.11 中得到的含有 p 个滞后项的模型的误差协方差行列式，而 m 是向量自回归模型中内生变量的个数。计算出的值越小，模型的拟合度就越高，因为在方程 3.14 和方程 3.16 中的前一项是方程 3.12 中似然函数的误差协方差行列式，而后一项是对模型参数数量的惩罚因子，含有更多参数的模型具有更大的惩罚因子。所以对滞后项个数 p 的选择是基于最小的信息准则计算值的。

需要注意的是，每一个被加入 AIC 或者 BIC 计算信息检验的滞后值都会将拟合标准的惩罚因子提高 m^2。增加一个滞后项所带来的变化将小于含有较少滞后项模型的惩罚因子的均值，因为拟合度之间的差别不足以抵消加入更多参数而为模型带来的损失。[14]

一般来说，我们用似然比检验和信息准则检验一起决定滞后时长。因为对向量自回归模型进行最小二乘估计是渐进一致的——模型估计的正确性会随着样本量的增加而提升——而加入更多滞后项造成的损失仅限于估计有效性的降低。但如果模型所包含的滞后项太少，则会遗漏一些动态关系。必须引起注意的是，信息准则检验可能会选择不同的滞后时长，这是由两个原因导致的：(1)信息准则的计算依赖于对 $p_{\max}$ 的选择，而 $p_{\max}$ 已经足够大(例如一年的月度数据)；(2)拟合标准可能会高估向量自回归的真实滞后顺序(Lutkepohl，2005：第 4 章)。

这里有必要提及有关滞后时长检验的最后两点。首先，所有的滞后时长检

验必须基于同一个样本。因为基于小样本的“小”模型所用的滞后项数量也是有限的，有一种尝试是让所有可能的观测值都充当滞后项，但这是不正确的。在寻找合适的滞后长度时，所有的卡方值和拟合标准都应该基于同一个样本进行计算。第二，我们可以设定滞后时长的上限，并且对含有一个滞后项到含有这个上限数量的滞后项的所有模型进行比较。对滞后时长进行选择的目的是用最简化的模型来解释尽可能多的动态关系。

检验残差的序列相关

当残差项不存在时间序列相关时，向量自回归模型的估计结果是十分稳健的。从这个意义上说，我们需要对向量自回归模型的残差项进行时间序列的检验，从而确定不同滞后时期的残差是不相关的。向量自回归模型在设计上允许在 Σ 或者 A_0 矩阵中，残差和同一时期的自变量之间存在相关关系。本章所关注的是，不同时间段的残差是否存在序列相关。

目前进行残差序列相关检验的几个方法都是通过估计一元自回归整合移动平均模型的残差来确定动态关系是否被正确地设定。普遍运用的是以下几种方法，我们将按其复杂程度逐一介绍。第一个方法是对所有变量和残差逐一进行散点图描述。第二个方法包括不同滞后时期的变量自相关和互相关散点图。第三个方法是运用混合统计量。最后一个方法是对滞后序列的检验，在之前已经讨论过。

在对残差中可能存在的序列相关进行检验时，第一个可用的方法就是图形法，即对各时段的残差绘制散点图。这一方法的问题在于，它在很大程度上依赖于分析者对序列相关的识别能力。但我们经常会遇到质量较差的数据，在这种情况下，运用这一方法的难度就加大了。有时序列相关模式是同时发生在几个变量之间。

第二个方法比简单的图形法有所进步。该方法是在每个时段都进行自相关和互相关函数的计算，从而检验是否存在残差的自相关。自相关函数是检验一个变量（在这里是残差）与其自身的过去值之间是否存在相关关系；互相关函数则是针对一个变量的残差是否和其他变量在过去时段的残差有相关关系。

这些相关关系可以被绘制成标准自相关函数的图形。运用散点图进行假设检验的方法在这里也同样适用。

第三种方法是混数检验。这种方法更为正式并被用来确定残差项是不是白噪音。在混数检验中，Box-Ljung 和 Q 统计量是一元时间序列建模过程中常见的，其他不同的统计量常见于多元时间序列模型中。这些检验是从一个给定滞后时长的拟合后的向量自回归中计算残差项的序列相关的。对每个检验来说，虚无假设是残差估计值与 h 个滞后项不相关，或者残差估计值与方程 3.7 中的 $T \times m$ 残差矩阵不存在序列相关。虚无假设和备择假设的数学表达式如下：

$$H_0: E[e_t', e_{t-i}] = 0 \ (i = 1, \cdots, h > p)$$

$$H_1: E[e_t', e_{t-i}] \neq 0 \ (\text{对某些 } i = 1, \cdots, h > p)$$

表达式 $E[e_t', e_{t-i}]$ 是在 t 和 $t-i$ 两个时段的残差的协方差。为了检验这一假设，我们为滞后项 h 建立一个多元 Q 统计量：

$$Q_h = T\sum_{j=1}^{T} \mathrm{tr}(\hat{\Gamma}_j' \hat{\Gamma}_0^{-1\prime} \hat{\Gamma}_j \hat{\Gamma}_0^{-1})$$

其中，tr( )是矩阵的迹[15]，$\hat{\Gamma}_j$ 是 t 和 $t-j$ 两个时段的残差协方差矩阵，h 是虚无假设所检验序列相关的滞后时长。残差的样本协方差 $\hat{\Gamma}_i$ 可以通过下列表达式来计算：

$$\hat{\Gamma}_i = T^{-1} \sum_{t=i+1}^{T} \hat{e}_t' \hat{e}_{t-i}$$

其中，e_t 是向量自回归模型的残差。如果模型的滞后时长为 0，那么协方差和残差的协方差矩阵就相同。因此，$\Gamma_0 = T^{-1} \sum_{t=1}^{T} \hat{e}_t' \hat{e}_t = \hat{\Sigma}$。统计量 Q_h 的分布具有渐进性，且服从自由度为 $m^2(h-p)$ 的 χ^2 分布。

对一元时间序列模型进行检验时，Q 统计量在遇到小样本时的效果较差。解决方案是在对从 1 到 h 个滞后值的残差进行序列相关检验时，修改 Q 或者 Box-Ljung 模式的混数统计量。修改后的 Q 统计量的表达式如下：

$$Q_h^* = T^2 \sum_{j=1}^{T} \frac{1}{T-j} \mathrm{tr}(\hat{\Gamma}_j' \hat{\Gamma}_0^{-1\prime} \hat{\Gamma}_j \hat{\Gamma}_0^{-1})$$

该统计量服从自由度为$m^2(h-p)$的χ^2分布。该式中的自由度是经过$T/(T-j)$因子纠正过的，用来表示对从1到j滞后项的相关关系进行估计。

在这两个检验中，对h的选择至关重要，因为如果在滞后项h前没有序列相关，我们可能就无法拒绝原假设。但是可能在h后的滞后项之间存在序列相关。所以，研究者必须选定一个足以反映数据周期性的序列值h。

另一个替代性的检验是运用布雷施—戈弗雷拉格朗日乘子法来检验多元回归的序列相关。针对多元回归的检验逻辑是，将最小二乘估计模型的残差项回归到因变量的滞后值和残差的滞后值，进而检验这一非限定模型中滞后残差项的回归系数是否为0。在多元回归中，我们运用对y_t的向量自回归模型(p)来对另外两个向量自回归模型进行拟合。

拉格朗日乘子检验法有4个步骤。第一步是对非限定的人为设定向量自回归模型进行估计。在这里，我们允许y_t的向量自回归模型的残差项存在序列相关的可能。这一步骤通过对下列向量自回归进行估计来完成：

$$e_t = y_{t-1}A_1 + \cdots + y_{t-p}A_p + e_{t-1}B_1 + \cdots + e_{t-h}B_h + u_t \qquad [3.17]$$

这一向量自回归是通过将残差项矩阵回归到y的第1到第p个滞后项的残差以及最初的模型中第1到第h个滞后项的残差。

第二步是对第二个人为设定向量自回归模型进行估计，这一模型是一个限定模型，其中$B_1 = \cdots = B_h = 0$：

$$e_t = y_{t-1}A_1 + \cdots + y_{t-p}A_p + u_t^R \qquad [3.18]$$

这一限定模型对应的虚无假设是最初的向量自回归模型的残差不存在序列相关。

第三步是为两个人为设定的向量自回归残差模型(方程3.17和方程3.18)构建一个残差协方差：

$$\tilde{\Sigma}_e = T^{-1}\sum_{t=1}^{T}\hat{u}'_t\hat{u}_t$$

$$\tilde{\Sigma}_R = T^{-1}\sum_{t=1}^{T}\hat{u}_t^{R\prime}u_t^R$$

其中，第一个残差协方差矩阵是从非限定人为设定回归中估计出来的，而第二

个残差协方差矩阵从第二个限定性的人为设定回归中得出。

最后一步是利用 χ^2 拉格朗日乘子检验统计量来判定向量自回归残差中是否存在序列相关。统计量的计算式如下：

$$\mathrm{LM} = T[m - \mathrm{tr}(\tilde{\Sigma}_e \tilde{\Sigma}_R^{-1})]$$

表达式中的 m 是系统中内生变量的个数，tr( )是矩阵的迹。统计量服从 χ^2 分布，自由度为 hm^2，即不存在残差序列相关的虚无假设下限定模型的参数个数。

最后一个也可能是最典型的确定向量自回归模型是否具有白噪音残差的方法是对模型进行过度拟合。具体做法是拟合一系列向量自回归模型并且检验具有 p^* 个滞后项的限定模型与具有 p^*+1 个滞后项的模型没有差别。这一检验的过程与先前描述的设定滞后长度的分析方法一样，即确保在残差具有白噪音性质的前提下找出最简化的模型。

总的来说，所有对残差进行序列相关的检验在样本容量增加时趋于相等。

如果这些检验方法得出了残差具有序列相关的结论，我们就必须往向量自回归模型中加入更多的滞后项。这些滞后项可以加入每一个方程，而非只显示在序列相关的变量所在的方程中。请记住，向量自回归方法中最典型的检验是过度拟合检验。这与之前概括的确定模型合适的滞后时长的方法一样。在大多数情况下，向量自回归模型的报告中都包含对滞后时长的选择和加入不同滞后时长的结果，以此表明对动态关系设定的稳健性。

格兰杰因果关系

在一个多元时间序列的非限定向量自回归模型中评估变量之间的关系时，涉及一个变量取值的重要问题，这一问题可以通过下列方式来表述：(1)在动态方程系统中，变量 Y_t 的哪一个值预测了 Z_t 的值？(2)在时间序列模型中，变量 Y_t 对于 Z_t 来说是内生的么？(3)变量 Y_t 对 Z_t 的未来值的关系是线性的么？这 3 个问题在多元时间序列模型中评价变量关系被认为是等同的，而问题的答案基于对时间序列模型中格兰杰因果关系的确定（详见 Granger，1969；Sims，1972）。

为了将这一概念进行精确的定义，我们用下列两个变量(Y_t, Z_t)的二元向量自回归模型来举例：

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{i=1}^{p} \beta_i Z_{t-i} + \varepsilon_{1t} \quad [3.19]$$

$$Z_t = \beta_0 + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{i=1}^{p} \delta_i Z_{t-i} + \varepsilon_{2t} \quad [3.20]$$

在这一方程系统中，格兰杰因果关系可以被定义为，对于线性模型，如果 Y_t 的过去值比 Z_t 的过去值更好地预测了 Z_t 的当下值，那么可以说，Y_t 格兰杰导致 Z_t。

上述定义的反命题依然为真。对于方程 3.19 和方程 3.20 来说，如果 Z_t 格兰杰导致 Y_t，那么在 Y_t 表达式中，Z_t 过去值的系数就不为 0，或者说 $\beta_i \neq 0$，$i = 1, 2, \cdots, p$。同样，如果 Y_t 格兰杰导致 Z_t，那么在 Z_t 表达式中，Y_t 过去值的系数就不为 0，或者说 $\gamma_i \neq 0$，$i = 1, 2, \cdots, p$。

格兰杰因果检验的一个途径是评估一个模型中的变量的过去值 $Y_{t-1}, \cdots, Y_{t-p}$ 是否预测了另一个变量 Z_t。最正式的检验方法就是通过对“可能的原因变量并没有导致结果变量的变化”这一虚无假设进行 F 检验或者 χ^2 检验。这一虚无假设被称为“非因果假设”，数学表达式如下：

H_0：格兰杰非因果　Z_t 不能预测 Y_t，如果 $\beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_A：格兰杰因果　Z_t 能够预测 Y_t，如果 $\beta_1 \neq 0$, $\beta_2 \neq 0$，或者 $\beta_p \neq 0$

请注意，备择假设是任何系数都不为 0。这一假设检验可以通过似然比检验或者 F 检验来完成。F 检验相对简单，需要以下两个回归模型[16]：

$$\text{模型 1(非限定)：} Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{i=1}^{p} \beta_i Z_{t-i} + \varepsilon_{1t}$$

$$\text{模型 2(限定)：} Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + u_{1t}$$

进行检验时，我们需要知道总方差(RSS)：

$$\text{RSS}_{\text{非限定}} = \sum_{t=1}^{T} \varepsilon_{1t}^2$$

$$\text{RSS}_{\text{限定}} = \sum_{t=1}^{T} u_{1t}^2$$

检验时，通过比较总残差平方服从自由度为(p, $T-2p-1$)的 F 分布：

$$F(p,\ T-2p-1) \sim \frac{(\mathrm{RSS}_{\text{限定}} - \mathrm{RSS}_{\text{非限定}})/p}{\mathrm{RSS}_{\text{非限定}}/(T-2p-1)}$$

如果这一 F 统计量的值大于所选定的显著性水平的临界值，我们就拒绝虚无假设，即 Z_t 对 Y_t 的效应为 0，得出 Z_t 格兰杰导致 Y_t 的结论。需要注意的是，同样的检验可以被用于检验 Y_t 和 Z_t 之间的格兰杰因果关系，即检验方程 3.20 中的 $\gamma_1 = \gamma_2 = \cdots = \gamma_p = 0$。

χ^2 检验也常被用于评估格兰杰因果关系。因为当方程中存在大量的变量和滞后项时，F 统计量会因为分子和分母的自由度趋近相同而失效，这就导致利用 F 统计量检验虚无假设会产生偏误。χ^2 检验是运用似然比或者沃德检验，其虚无假设与 F 检验时相同，即所有的系数都同时为 0。

两个检验方法具有渐进的相等性，但是 F 检验更容易操作，且对于两个变量间的因果假设的检验效果较好。

解释格兰杰因果

对格兰杰因果的诸多批评之一是，它并不满足“哲学意义”上的因果定义。这一批评是正确的，因为对于因果关系的标准定义需要找到时间上一致的(一个变量的变化发生在另一个变量变化之前)、统计上显著的非虚假相关(例如，Hamilton，1994:302—309)。格兰杰因果显然满足前两个条件，但却不一定满足第三个。在前面所讨论的 F 检验和 χ^2 检验只能检验非零滞后值的系数，所以该检验无法评估关系的方向。为了确定变量间的相关关系并不是虚假的，我们需要理论作为指导。

在运用格兰杰因果评估变量之间的时间关系时，我们需要考虑的第二个问题是同期相关。请思考如下例子：假设格兰杰相关假设检验显示，我们不能拒绝非因果性的虚无假设，这就意味着一个变量的过去值不能预测另一些变量的当下值。但是在这样一个系统中，变量可能具有很强的自回归并能够很好地自我预测。如果在多元时间序列模型中存在很高的同期相关性，那么创新或者外部冲击将会是相关的。因此，尽管“因果关系”中的“过去值”不能是现在的，但

是时间序列可能存在高度的同期相关性——两个时间序列可能受到普遍相关的创新的影响。因此,格兰杰非因果性的证据并不能说明系统中的时间序列是不相关的。换句话说,变量的过去值之间并不具备预测性。具体例子请见威廉姆斯和麦金尼斯的著作(Williams & McGinnis, 1988)。

第三个需要注意的问题是模型的设定。这包含两个潜在问题:错误的滞后时长设定和忽略与格兰杰原因变量有关的变量。假设在向量自回归模型中,时长设定是不正确的,那么就有以下两种情况:太多滞后项或太少滞后项。如果一个模型包含了太多的滞后项,估计结果可能将是无效的,但依然是无偏的(仅当模型是线性回归时)。因此,假设检验也将是无偏的但是无效的。所以我们很可能在应该拒绝虚无假设时无法拒绝它。我们再来考虑滞后项太少的情况。在这一情况下,向量自回归估计是有偏且无效的,这种情况就像线性回归中的忽略变量。忽略滞后项可能意味着未能把一些动态关系考虑到模型中,这将会导致残差项的序列相关。统计量检验的结果将会"太好"以至于让我们很容易拒绝虚无假设。因此,在向量自回归模型的设定中忽略变量和动态关系将会导致我们找到本不存在的因果相关。

另一个可能的设定问题是忽略一个和格兰杰原因变量相关的变量或者方程,从而影响系统中的所有变量。例如,我们假设一个含有两个变量的系统,变量是内生的且统计检验支持格兰杰非因果的虚无假设。假设可能存在第三个变量,如果将其纳入模型中则会推翻虚无假设,但这一变量几乎不可能存在。因为利特曼和威斯(1985)证明,忽略这样一个变量是不可能推翻内生性或非因果性的证据的(Litterman & Weiss, 1985)。这是因为在前两个变量中纳入一个能够推翻非因果性假设的变量时,这个新纳入的变量必须能够完美地排除前两个变量的所有滞后项所得到的内生性结果,但这几乎是不可能的。

最后一个可能影响格兰杰因果推论的问题是变量中单位根和随机趋势的出现。如果一个或多个变量含有一个单位根,那么对模型参数的统计检验将是一个非标准分布(Sims et al., 1990)。这些检验具有非标准分布的原因和将单位根变量回归到一个稳定变量时遇到问题是相同的。格兰杰和纽博德以及菲利普证明,如果把一个单位根变量回归到与其不相关的另一个变量,但由于单位根

变量存在一定的趋势，所以回归系数将是不正确但统计上显著的(Granger & Newbold, 1974; Philips, 1986)。因此，对格兰杰因果的评估可能因单位根变量的出现而严重地受影响。

从频数概率的角度看，当单位根出现时，对格兰杰因果的检验将会因干扰参数而变得复杂。一些研究者已经开始改良格兰杰因果检验，以便在单位根出现时提供正确的统计检验结果(Dolado & Lutkepohl, 1996; Lutkepohl & Reimers, 1992a; Zapata & Rambaldi, 1997)。单位根的出现所引发的最基本问题就是模型无法包含一些动态关系，在这种情况下，标准的 F 检验和 χ^2 检验都是不正确的。许多研究者运用各种方式来拓展对格兰杰因果关系的检验，这些方法主要聚焦于模型中单位根的数量。如果模型中含有 m 个变量，那么最多只能含有 $d \leqslant m$ 个单位根。这一方法被用来对含有 $p+d$ 个滞后项的模型进行估计，并且基于前 p 个滞后项来对模型进行格兰杰因果关系检验。另一个方法是运用一阶差数据，这样在检验时，统计量就不会出现上文所描述的非标准分布的现象。

尽管格兰杰因果关系存在一些问题，但它依然是建立和检验模型的有效工具。需要注意的一点是，格兰杰因果只有助于对预测性变量的估计，而对于多元时间序列模型中的结构参数的统计推断，格兰杰因果并没有什么效果。因为向量自回归模型是运用最小二乘法估计的，尽管在格兰杰非因果的限制下，估计值依然是一致的。此外，只要模型的滞后项个数足以确定残差项具有白噪音性质，那么检验就是有效的。这一点极其重要，如果我们通过检验认定模型是格兰杰非因果的，我们接下来必须检查格兰杰因果关系是否受到滞后项个数的影响。

对向量自回归模型其他限定条件的检验

因为向量自回归模型建立在多元回归模型的标准步骤上，所以假设检验就可以通过标准的回归分析中的 F 检验和 χ^2 检验来完成(除非模型中含有非稳定变量)。这包括对单个系数或一组系数的显著性检验。

向量自回归模型有一个特殊的假设检验，即组外生性检验。这一检验是对

单个方程中的一组特定的滞后项或者特定变量的多个滞后项施加多个限定条件。通过似然比检验，我们可以比较多个方程中的一组系数是否存在不同。这相当于运用 χ^2 检验来评估一组方程的系数是否为 0。这一检验的基本思想是，如果虚无假设成立，那么限定模型和非限定模型之间的似然比就会较小。但如果虚无假设不成立，那么这一似然比就足以让我们拒绝原假设。

冲击反应和移动平均反应分析

前述的所有检验对我们理解多方程时间序列系统中变量之间的关系都很有帮助。但是我们最终的目的是运用向量自回归模型来描述序列之间的动态行为，其中最常用的方法是冲击反应和移动平均反应分析。

移动平均反应与时间序列模型的动态乘子分析法类似(例如自回归整合移动平均模型)。在对向量自回归模型进行估计时，每个方程的残差项或者外生冲击都是随机影响。因此，残差项对系统中的方程来说是一个外生冲击。移动平均分析就是解释这些外生冲击对模型的向量移动平均表达形式的影响。

向量自回归模型建立在向量移动平均表达形式的基础上。每个稳健的有限阶自回归时间序列模型都可以被写成一个无限滞后移动平均时间序列模型(Wold, 1954)。这一性质的有用之处在于，它可以将模型围绕其平衡值再次集中，从而观察外生冲击对时间序列的影响。这些外生冲击的动态性建立在向量移动平均表达式的基础上。

通过将自回归过程的滞后多项式的因子化，我们可以将向量自回归模型改写成一个向量移动平均表达式。这里的滞后多项式是针对一组时间序列的滞后项建立的矩阵函数。例如，我们可以将一个含有两个滞后项的向量自回归模型写成如下多项式形式：

$$y_t = c + y_{t-1}B_1 + y_{t-2}B_2 + e_t$$

$$y_t = c + y_t(B_1L + B_2L^2) + e_t$$

其中，L^k 是一个滞后算子，变量与其相乘就会被转化为 k 个阶段之前的值，即 $L^k y_t = y_{t-k}$。有了这个算子，我们就可以将方程系统的滞后多项式写成一个更紧

凑简洁的形式。我们接下来会看到，误差来自其他变量如何通过自回归系数矩阵来对模型产生影响。

运用上述形式，我们将一个含有 p 个滞后项的向量自回归模型改写成向量移动平均表达式：

$$
\begin{aligned}
y_t &= c + y_{t-1}B_1 + y_{t-2}B_2 + \cdots + y_{t-p}B_p + e_t \\
y_t - y_{t-1}B_1 - y_{t-2}B_2 - \cdots - y_{t-p}B_p &= c + e_t \\
y_t(I - B_1L - B_2L^2 - \cdots - B_pL^p) &= c + e_t \\
y_t &= (c + e_t)(I - B_1L - B_2L^2 - \cdots - B_pL^p) - 1 \\
y_t - d &= e_t(I + C_1L + C_2L^2 + \cdots)
\end{aligned}
\tag{3.21}
$$

其中，d 是向量自回归模型的常数项，是由 c 除以自回归滞后多项式而得到的。这个推导过程的最后一步是为一个无限滞后向量移动平均表达式生成一组系数，从而使其与有限阶含有 p 个滞后项的向量自回归模型相互匹配。具体做法是让自回归系数的多项式与一组相应的移动平均系数相等：

$$
(I - B_1L - B_2L^2 - \cdots - B_pL^p)^{-1} = (I + C_1L + C_2L^2 + \cdots)
$$

$$
(I - B_1L - B_2L^2 - \cdots - B_pL^p)(I + C_1L + C_2L^2 + \cdots) = I
$$

上式的左边是 p 阶 m 维矩阵多项式的自回归表达式，右边则是无限阶 m 维移动平均多项式的表达式。

方程 3.21 中的移动平均反应系数 C_i 可以通过以下递推式的自回归系数算出：

$$
\begin{aligned}
C_1 &= B_1 \\
C_2 &= B_1C_1 + B_2 \\
C_3 &= B_1C_1 + B_2C_1 + B_3 \\
&\vdots \\
C_t &= B_1C_{t-1} + B_2C_{t-2} + \cdots + B_pC_{t-p}
\end{aligned}
\tag{3.22}
$$

对于任何 $j > p$，都有 $C_0 = I$ 并且 $B_j = 0$。模型的无限滞后移动平均表达形式与自回归表达形式所包含的动态关系是相同的。

我们之所以进行上述两种表达形式的转换，是因为后者简化了模型的动态关系结构，从而能让我们勾勒出外生冲击或创新对系统的影响。运用向量移动平均表达形式，我们可以分析每个方程中外生冲击通过 y_t 对 e_t 的影响。在这里，使用向量移动平均表达形式能让我们看到外生冲击与零均值的差别是如何随着时间变化并消失的。我们可以通过方程 3.21 来达成这一目的。该方程的右边是每一个时间序列在长期变化中的向量自回归平衡，而左边所描述的是这些时间序列是如何围绕这一平衡而变化的。

另一个解释向量移动平均表达形式的方法是将其视为一个方程中外生冲击所带来的影响。这可以写成一个设定外生冲击影响范围的矩阵导数。外生冲击变量 j 在时间 s 的变化所造成的 y 在 l 时间的变化可以被记为 $e_j(s)$，它是该外生变量的方程 3.21 中 y_t 的导数，或者写为：

$$\frac{\partial y_i(l+s)}{\partial e_j(s)} = c_{ij}(l) \qquad [3.23]$$

其中，$c_{ij}(l)$代表方程 3.21 和方程 3.22 中 C_l矩阵的第 i 行第 j 列。这一数值代表着方程 i 对于外生冲击 j 在时间 $s > l$ 所带来的影响的反应强度。请注意，这里是向量形式，所以在系统中，一个变量对于一个外生冲击只有一个反应值。因此，如果系统中有 m 个变量，那么将会有 m^2 个冲击反应（包括每个变量作为自身的外生冲击）。

对向量自回归的平均移动反应进行求导可以为模型中系数之间的动态关系提供 3 种解释。第一种是将其解释为外生冲击对系统的影响；第二种是将其解释为动态乘子效应的改变对模型中多个自变量的影响；最后一个解释如方程 3.23，将其视为一个方程的外生冲击对模型中自变量的边际效应。换句话说，我们在这里分析的是一个方程的外生冲击 e_t 对方程系统中每个自变量的边际效应。由于内生变量的改变具有动态性，所以我们必须观察它随时间的变化趋势。这 3 种解释都涉及对系统外生冲击的设定问题。在对方程 3.22 的推导过程中，我们假设变量的同期反应是不相关的，或者说 $C_0 = I$。但宽泛地说，同期的外生冲击之间并不是不相关的。因此，我们就需要对误差项是否相关作出假定。

这些假定有时可以通过理论的指导来作出，因为肯定有一些理由让我们假设在一个时间序列中的外生冲击与另一个是不相关的。但在另一些情况下，我们可能无法对这个外生过程作出任何先验的假设。另外，向量自回归模型中变量的序列效应和向量移动平均系数的分解也会带来影响。从这个意义上来说，解释向量自回归模型的系数似乎遇到了和结构方程模型一样的困难。但是，在向量自回归建模过程中，对于同期关系的设定实际上是在解释了数据的动态关系之后作出的。这说明我们所作出的设定并不是有偏的，也没有限制我们对动态关系的解释。

动态关系的设定所起的作用可以通过冲击反应矩阵来看。在 $m \times m$ 矩阵 C_l 中，在时间 l 下有 c_{ij} 个元素，那么 i 个方程(矩阵的行)在 l 时点对变量 j (矩阵的列)的冲击所作出的反应可以表示为：

$$C_0 = A_0^{-1},\ C_t = \sum_{j=1}^{t} C_{l-j} B_j (i = 1,\ 2, \cdots) \qquad [3.24]$$

这里，我们规定 $j > p$ 时，$B_j = 0$。方程在 0 时段对外生冲击反应可以由 A_0^{-1} 的同期相关获得。这一相关通过第二个递推方程中的自回归系数计算而逐渐传递到下 s 个时段，结果是每一个时段都有 s 个矩阵来计算移动平均效应。因此，移动平均效应的计算依赖于对同期残差项的最初设定。

基于数据和理论，我们可以将同期相关的计算归为以下几种方式：第一，如果我们对于外生冲击影响系统变量的顺序一无所知，那么就可以先通过对误差协方差矩阵 $\hat{\Sigma}$ 进行考利斯基分解(Cholesky)来计算 A_0^{-1}。我们可以发现，矩阵分解 $A_0^{-1\prime} A_0 = \hat{\Sigma}$。然后，我们寻找矩阵 $\hat{\Sigma}$ 的平方根，它是一个下三角矩阵(在矩阵 $A_0^{-1\prime}$ 中)。这一过程也被称为“残差的正交化”。第二个方法是通过对 $A_0^{-1\prime}$ 的估计对矩阵 $\hat{\Sigma}$ 进行因子化分解。如果没有很强的理论指导，这一方法通常很难实现。最后一个方法是直接寻找系统中每个方程的矩阵 A_0。

我们在这里将介绍最为普遍的方法，即对误差协方差矩阵 $\hat{\Sigma}$ 进行考利斯基分解来计算 A_0^{-1}。在这种情况下，变量的序列效应是十分关键的。如果残差项中的相关十分弱，那么变量的序列在计算方程的反应时并不是一个主要的因素。但是，当时间序列之间是高度相关的时候，那么变量的序列效应将会影响

对结果的解释。因此,我们就需要评估变量的多种序列组合来对估计的稳健性进行检查。

我们要作出一个很重要的提醒,即无论选择什么样的方法来计算残差项的同期相关,我们都必须对同期的外生冲击作出设定。如果系统中的变量对外生冲击的反应是彼此不相关的,那么对同期关系的设定就无关紧要。但如果这些反应是高度相关的,那么我们就必须考虑以下几种情况:首先,如果向量自回归模型中只有 m 个自变量的子集之间存在高度的同期相关,那么就先将这些变量一起放入考利斯基序列中。虽然我们不知道这些变量如何互相影响,但是我们可以确定它们作为一个整体对系统的影响。其次,我们也可以运用多元分析来看系统对外生变量的反应是如何受到不同的同期序列设定的影响的。最后,我们可以考虑另外的因子分解方法来进行检查。

在识别了残差的相关关系或者同期关系之后,对冲击反应的计算可以通过一组 s 个 $m\times m$ 矩阵完成。具体做法是,先将矩阵 C 中的元素在 s 个滞后项的 m^2 个时间序列中重新排列,然后再用图形展示 s 个滞后项的反应的一组 $m\times m$ 矩阵。这样一来,m^2 个序列就被置入一个 $m\times m$ 维的图中。此图所反映的就是我们对于 A_0^{-1} 矩阵的识别假设,即在矩阵 A_0 的行中使用同样排列顺序的变量。具体的表述方法我们将在下一章中呈现。

误差范围冲击反应

冲击反应被用来追踪向量自回归模型中的方程如何对一组设定的外生冲击作出反应。由于在广义稳定自回归过程中存在移动平均效应,我们可以认为,这些对外生冲击的反应最终会收敛到 0。并且,对冲击的设定一般假定冲击的强度是向量自回归模型残差项的一个标准差的大小。这一反应是一个随时间变化的函数。

目前为止,我们的讨论都聚焦于如何估计和计算冲击反应的均值,接下来我们着眼于外生冲击在影响变量 j 时,方程 i 的冲击反应的移动平均过程是否随时间而变化。如果反应的置信区间不包含 0,或者在从 0 到 s 的多个时段中都不为 0,那么该反应可以被认为是统计上显著的。相反,如果反应的置信区间在

时间的变化过程中包含 0，那我们就没有足够的统计证据来证明该反应不为 0。[17]

由于冲击反应的估计是基于自回归系数及其移动平均表达形式的，所以可以被认为是随机的。它反映了一个变量对于外生冲击的动态反应，所以也具有均值和方差。均值的计算是把对方程系统的冲击序列效应视为自回归系数估计的函数。若对冲击反应的置信区间进行估计，我们就必须确定自回归系数变化的两个来源，即系数本身和残差的协方差阵。

相较而言，对冲击反应的方差的估计和推导困难一些。困难存在于两个方面：首先，计算方差过程本身就十分复杂，并且要基于对现实数据的一些假设，而这些假设有可能是不正确的。因为对向量自回归模型参数的最大似然估计只有在样本量趋于无限大时才服从正态分布，这就意味着在有限的样本量下，向量自回归系数只能是近似正态的（Lutkepohl, 1990）。但是冲击反应系数是这些近似正态参数的非线性函数（实际上是矩阵多项式）。因此，我们并不能保证这些反应系数是正态分布且仅用均值和方差就可以进行描述的。第二个难点在于，对外生冲击的反应可能存在自身的序列相关。这是向量自回归模型自身的结构导致的。该模型的基础是对变量的滞后项如何影响当下值的设定。如果滞后变量的值由两部分组成，即一个固定项和一个由外生冲击或创新组成随机项，那么正如向量移动平均表达式所呈现的，变量后续的创新依赖于较早时段的创新。尽管我们在估计过程中假设创新之间或者残差之间是不相关的，但是在对这些冲击反应的动态分析中，向量自回归模型将其设定为相关的。这就意味着，我们在求方差的过程中，必须考虑潜在冲击反应之间的序列相关。

研究者已经提出一些方法来推进对不确定性或者误差范围的测量，从而对方程 3.23 中冲击反应的置信区间作出描述（Runkle, 1987）。但当冲击的时间范围增加时，这些分析推导方法的效果往往很差。基利恩提出了一个基于冲击反应置信区间的小样本“二重自助法”（Kilian, 1998）。这一方法能够减少向量自回归模型系数最初的估计偏误，但却无法解决非高斯、非现象反应之间高度相关的问题（详见 Sims & Zha, 1999:1125—1127）。

计算误差范围的标准方法依靠建构下列区间：

$$\hat{c}_{ij}(t) \pm \delta_{it}(t) \tag{3.25}$$

其中，$\delta_{ij}(t)$是在时间 t 上、标记为 $c_{ij}(t)$和 $100(1-\alpha)$ 的置信度下定义冲击反应置信区间上下限的函数。通过绘制$\hat{c}_{ij}(t)-\delta_{it}(t)$ 、$\hat{c}_{ij}(t)$和 $\hat{c}_{ij}(t)+\delta_{it}(t)$ 3 个函数随时间 t 的散点图，我们可以将它们图形化。这就是所谓的误差范围“连点图”，并可以用多种统计软件输出（例如 RATS、Eviews 和 Stata）。利用这种方法计算冲击反应方差时，我们假设冲击反应之间不存在跨时段的相关关系，即函数 $\delta_{ij}(\mathrm{t})$与 $\delta_{ij}(t+k)$ 之间是不相关的。这个假设意味着，两个函数除了通过自回归参数的变化产生联系外，未来冲击反应的方差不依赖于过去的冲击反应方差。

计算函数 $\delta_{ij}(t)$的误差带有以下几种方法：最常用的方法是模拟一个冲击的样本并概括这个样本的性质。具体做法是从向量自回归系数的后设分布中计算蒙特卡洛样本。[18]在这一方法中，后设样本是通过向量自回归系数的渐进分布建构出来的。基于方程 3.7 和方程 3.10，这些系数服从均值为 $B \sim MVN(\hat{B},\ \Sigma \otimes (X'X)^{-1})$的多元正态分布。将残差协方差的后设估计进行转置，可以得到 Σ^{-1}，服从维舍特分布（Wishart），即 $\Sigma^{-1} \sim \mathrm{Wishart}(\mathrm{S},\mathrm{T})$，其中 $\bar{\mathrm{S}}$ 是 $\hat{\Sigma}^{-1}$ 的样本估计值（详见 Zellner，1971）。利用上述条件，我们可以按照如下步骤来建构一个向量自回归冲击反应的样本：(1)将维舍特分布进行转置从而得到 Σ^{-1}；(2)运用上一步的 Σ 从 $B \sim MVN(\hat{B},\ \Sigma \otimes (X'X)^{-1})$中抽出自回归系数；(3)通过方程 3.21 和前两步中抽出的自回归系数来计算冲击反应；(4)见第三步的反应估计值储存；(5)重复前四步 N 次，直到 N 足够提供一个近乎精确的冲击反应值。

概括 N 组 $m \times m \times s$ 反应将会产生冲击反应的估计方差。常用的样本总量 N 大约为 1000 到 5000，取决于我们所需要的精确度。我们可以利用更大的 N 来进行稳健性检验。

从这个冲击反应的样本中，我们可以计算一个近似正态的 $\hat{c}_{ij}$ 的均值：

$$\hat{c}_{ij}(t) \pm z_{\alpha}\sigma_{ij}(t)$$

其中，z_α 是正态概率密度分位数，$\sigma_{ij}(t)$是 $c_{ij}(t)$的标准差，而 $c_{ij}(t)$是 t 时段变量 i 对外生冲击 j 的反应(在 68%的误差置信度下，$z_\alpha = 1$；在 95%的误差置信度下，$z_\alpha = 1.96$)。可以看出，我们假设冲击反应在小样本中也是服从正态分布的。因此，围绕冲击反应均值的误差范围将是系统性的。

尽管高斯近似方法被用于先前一些主要的时间序列统计软件(如 RATS、Eviews 和 Stata)，但是目前也出现了一种替代性的方法(主要在 RATS 软件中比较常用)。该方法首先计算每个时点的 $c_{ij}(t)$的样本，然后从该样本中计算冲击反应的经验分位数和经验百分位数。在这种情况下，我们对后设区间的估计就应该基于最高的后设密度或者分位数，估计如下：

$$[c_{ij.\alpha/2}(t),\ c_{ij.(1-\alpha)/2}(t)]$$

其中，下标 $(1-\alpha)/2$ 和 $\alpha/2$ 表示置信区间的上下限(0.05、0.1、0.16 等)。

在解释移动平均反应激起误差范围时，我们要尤为注意。在向量自回归模型中，我们只估计了有限的参数。当这些参数被用于产生一组 $s > p$ 的更高维度的冲击反应时，对变量反应的推测将会十分不准确，也会导致我们整体预测的不准确。冲击反应的置信区间和误差范围也会因此而随时间范围发生指数级的增长。这是因为在向量自回归模型中，变量的不确定性和预测误差将会随着较长的冲击反应范围而增大。由于误差是随着时间而增长的，所以我们在解释后续阶段的冲击反应时就需要格外小心。

解释冲击反应时，需要注意的第二个问题是评估模型的稳定性。移动平均反应过程中的冲击反应不会迅速膨胀，所以在稳定的系统中，冲击反应会逐渐减小到 0。但是，在小样本中，我们可能会看到冲击反应的迅速膨胀。小样本(非限定的)向量自回归模型也许会出现一些不稳定，但是这种不稳定的误差反应是有偏的，并且其误差范围可能会很大。

面对上述问题，我们提出一些解释移动平均反应及其误差范围的准则。第一，请记住，反应来自系统的简化形式。因此，在第一个时段之后的冲击反应不仅包含对外生冲击的反应，还有一部分是来自系统内其他方程的反馈。因此，冲击反应同时呈现外生冲击的直接和间接效应。如果要评估其中直接效应的

大小，就必须借助 F 检验和格兰杰因果。第二，在运用移动平均反应过程对反应的最初阶段完成预测后，反应的标准误将会相对较快地增长。因此，误差范围中的一部分反映的是创新本身的不确定性。第三，如何决定冲击反应非零的可能性有多大？西姆斯认为，向量自回归模型为系统的不确定性提供了很好的估计，但是估计结果并不具有渐进性(Sims，1987)。因此，对经典的小样本假设的解释可能并不稳健。从这个方面来说，我们应该运用贝叶斯误差范围。

创新的计算和预测误差方差的分解

在解释多元时间模型中相互关联的动态关系时，一个常见的方法是创新计算和对预测误差方差的分解。这一方法运用内生变量在不同时段的变化来估计方程系统中每个内生变量的变化总量。

我们有两种方法来解释和理解创新的计算：第一种是分析变量与预测路径的离差。假设在多元时间序列模型中的变量遵循估计系数的预测路径，这些预测值包括两个组成部分，即预测出的变量路径和预期之外的创新和外生冲击。对于每个向量自回归系统，我们可以计算在因变量的预测路径中有多少方差来自因变量自身的过去值，又有多少来自方程中其他变量的过去值。这种对方差的分解就是对变量随时间变化的总方差进行计算和分解。第二种方法是将创新的计算当做广义的方差分析。因为我们处理的是多元时间序列，而方差分析则是针对在同一个时间段内，内生变量的方差有多少是被其他变量所揭示的。创新计算可以被看做跨时段的方差分析。

创新计算的具体做法是将每个方程中变量的方差的总量分解为其他每一个变量所解释的方差。这就等于探索每个变量的预测误差或者创新是如何影响系统的方程及其方差的。在这里，创新是时间序列中无法预期的部分，因此我们必须分析模型中预测误差的来源。

创新计算可以通过计算向量自回归模型中的预测误差的方差来完成。向量自回归的向量移动平均表达式(方程 3.21)可以被用于完成这一计算。利用方程 3.21，我们可以计算向量自回归系统在 s 时段的预测误差：

$$y_{t+s}-\hat{y}_{t+s}=e_{t+s}+C_1e_{t+s-1}+C_2e_{t+s-2}+\cdots+C_{s-1}e_{t+1} \qquad [3.26]$$

方程的左边是在 $t+s$ 时段，内生变量的观测值和估计值之间的差别。右边是预测误差从 $T=s$ 时段到 $s-1$ 时段的向量移动平均表达式。这说明模型中当下的创新和预测误差是过去创新的函数。系数 C_i 是移动平均系数。这一方程所表达的意思是，模型中的创新（冲击或残差）是其在向量移动平均表达形式中自身过去值的函数。

方程 3.26 中的预测误差的方差可以写成如下形式：

$$\begin{aligned}V(y_{t+s}-\hat{y}_{t+s})&=E[(y_{t+s}-\hat{y}_{t+s})'(y_{t+s}-\hat{y}_{t+s})]\\&=\Sigma+C_1\Sigma C_1'+C_2\Sigma C_2'+\cdots+C_{s-1}\Sigma C_{s-1}'\end{aligned} \qquad [3.27]$$

其中，$\Sigma=E[e_t'e_t]$ 是 t 时段预测误差的协方差。

在方程 3.27 中将预测误差按照时段进行分解的做法可以看做将系统变量的相对影响进行分割。这是描述系统中不同变量相对重要性的创新计算。尽管这一计算过程可以告诉我们每个变量的方差有多少可以被自身的过去值解释，但我们还想知道如何把第 j 个变量过去值的影响和其他变量的影响相互分离。这与方差分析相似，因为我们想知道的是一个变量的方差有多少是其自身的创新带来的，而又有多少是其他变量在不同时间段的创新中带来的。

完成这一方差分析最常见的方法是将方程 3.27 进行正交化。我们在冲击反应的章节已经提及，正交化主要有以下两个目的：一个是将外生冲击或创新的方差标准化；另一个是在预测创新之间建立同期关系，即这些同期误差是如何相互关联的。后一个目的对于我们来说更为重要，因为它关乎对预测创新的线性组合作出什么样的设定。需要指出的是，我们所做的是对向量自回归分析中变量的同期关系进行识别预设。

预测创新的正交过程可以写成下式：

$$e_t=u_tA_0^{-1}=u_{1t}a_1+u_{2t}a_2+\cdots+u_{mt}a_m \qquad [3.28]$$

其中，a_i 是残差的协方差 $\Sigma=A_0^{-1}{}'A_0^{-1}$ 的 i 列的分解。正如先前所提到的，这是对误差协方差矩阵进行考利斯基分解（是对矩阵平方根的广义化）。矩阵 A_0^{-1}

是对创新的同期相关进行计算。由于这是一个下三角矩阵，所以正交后的残差可以解释残差线性组合的特定模式。

我们可以运用这一正交法，按照时段和每个变量对预测误差的方差进行分解。用方程 3.28 中的正交后的残差协方差矩阵替换方程 3.27 中的残差协方差矩阵，那么方程 3.27 中的正交后的预测误差方差就可以写成：

$$V(y_{t+s}-\hat{y}_{t+s})=\sum_{i=1}^{m}a_i a_i'+C_1 a_i a_i' C_1' + C_2 a_i a_i' C_2' + \cdots + C_{s-1} a_i a_i' C_{s-1}' \qquad [3.29]$$

这相当于把移动平均表达式调整后放入正交后的残差。我们可以计算 s 个预测时段的每个方差矩阵。矩阵的第 i 行是方程的方差，第 j 列是被第 i 行的变量所解释的方差。因此，这 s 个预测方差矩阵让我们知道，每个变量的变异有多少来自其自身的创新，又有多少来自系统中其他变量的创新。

预测误差方差的分解结果一般是用交互表或者图形的方式来呈现。在表格中，列变量解释了行变量在 s 时段的预测误差方差的相应百分比，其中 s 是创新发生之后的步骤数量。一个严格意义上的完美的外生变量对于每个 s 来说，列变量的值都应该是 100。直观地看，这样的表格呈现了预测误差是如何在系统中进行反馈的。如果一个变量对于解释另一个变量的动态过程来说十分重要，那么前一个变量的预测误差就会导致后一个变量的预测误差。

预测误差分解之所以重要，是因为它有助于评估一个变量的变化是如何受其他变量影响的。这有助于我们确定格兰杰因果关系，并理解不同时间序列之间是如何相关的。创新计算有两个解释方法，一个是预测效果分析，另一个是方差分析，两者都对向量自回归系统的多元时间序列的方差作出动态的解释。但是，由于方差的分解是以百分比形式计算的，所以我们并不知道这一百分比对于原始变量的单位来说意味着什么。并且，这一方法也不能测量外生冲击影响方程系统时方差的不确定性。冲击反应分析和移动平均反应可以解决这一问题，因此我们在下文中将看到如何将创新计算和冲击反应分析运用于向量自回归模型中。

其他设定问题

在设定和估计向量自回归模型时，还存在一些其他的问题。第一个问题是模型中变量的测量和设定。这一过程中有关如何运用一阶差、稳定时间序列和数据转换(例如对数变换)等问题都需要被考虑。另外，还有一些其他因素能够影响向量自回归模型的推断，我们将在本章的最后一节逐一讨论。

差分方法能否用于趋势数据?

在一元自回归整合移动平均模型中，数据的转换对于分析来说极为重要。运用博克斯—詹金斯技术可以确定模型是否稳定以及残差是否具有白噪音性质。在向量自回归模型中，单位根的出现需要引起我们的重视，因为对参数估计和参数的分布都要求协方差是稳定的(Hamilton，1994)。确定协方差的稳定性需要知道模型外单位循环的多元滞后多项式的根。这一点是很难评估和检验的(尽管 RATS 等时间序列软件会提供一些常用的方法)。所以，分析者经常换用其他方法，即通过检查单位根和移动平均反应来看冲击是否消失或膨胀。含有单位根变量的模型可以直接用向量误差纠正模型或者其他模型来估计。需要注意的是，当含有单位根且其趋势没有在模型中被正确地设定时，向量自回归模型的估计结果是不可靠的(Hamilton，1994)。

上述内容暗示我们可以不必运用求差的方法来去除变量中的趋势性。实际上，向量自回归模型存在复杂的误差纠正机制，并能产生一个含有非稳定时间序列的稳定系统。在这种情况下，对数据中动态关系的推断就可能具有渐进性(Freeman，Williams，Houser & Kellstedt，1998；Sims et al.，1990)。

另外，求一阶差的方法还会去除时间序列中的一些长期效应，这些动态关系实际上是我们分析的主要目标，而求差的方法却排除了我们想要解释的大部分趋势和变化路径。

数据转化和清理

向量自回归模型中还有哪些常用的数据变化方式呢？一种常见的变换是趋势的对数形式或稳定变量的方差。运用数据的对数形式意味着，模型所表达的是数据的跨阶段增长率。这一解释来自下列增长率的渐进形式(在经济学中较为常见)：

$$\frac{y_t - y_{t-1}}{y_{t-1}} \approx \ln(y_t) - \ln(y_{t-1}) \qquad [3.30]$$

方程的左边是测量时间序列数据中跨阶段增长率的常见形式，而右边则是其渐进形式。在模型中，我们估计的实际上是右边的渐进形式，因为在向量自回归模型中，t 时段的内生变量是在方程的左边，而滞后值是在方程的右边。

运用自然对数变换能够让我们解释移动平均过程中，变量因为冲击而产生的百分比变化。这一方法对解释模型中百分比形式的变量(例如人口增长率)和水平变量(例如国民生产总值或美元计价的变量)十分有帮助。我们将在下文的例子中讨论这一数据转换及其解释。

另一个常见于时间序列分析的数据转换方法是对数据进行提前清理和过滤。在这种变换形式中，我们对数据进行特定的处理(如去除季节性)。但运用向量自回归时，我们要避免这类转换，这是因为清理和过滤数据会改变数据的动态性，因此影响对系统中时间序列的关系的估计。在一些情况下，清理数据甚至会彻底改变我们对动态关系所做的格兰杰因果评估(Sims，1972)。

向量自回归模型中的单位根和误差纠正

我们在第 2 章中曾经讨论过另一个建模策略，它是由伦敦经济学院的学者提出的，即误差纠正模型。该模型设定了变量在两个或多个时间序列中，如何通过各种短期或长期动态关系彼此相关。利用普通随机趋势来对数据进行建模，误差纠正模型可以用于稳定和非稳定数据。如果一个单位根或者随机趋势变量必须经过 d 次求差才能变为稳定数据，那么我们将其描述为 d 阶或 $I(d)$。

误差纠正模型能够为这类时间序列的趋势以及围绕这些趋势的多种动态关系提供清晰的描述。这一特殊模型的问题在于存在虚假回归的风险，即如果我们没有正确地设定数据的趋势，那么最终的因果推论就将是不正确的(Granger & Newbold, 1974)。

单位根数据的误差纠正表述

针对趋势性数据运用误差纠正模型实际上是格兰杰表述定理的结果，该定理说明，如果两个单位根变量具有同一趋势，那么两个时间序列就存在一个稳定的线性模型。这一定理的重要性包括以下 3 方面：第一，该定理意味着，当我们运用一个模型来估计两个或多个时间序列是否具有同样的趋势和动态关系时，避免了可能存在的虚假回归。第二，该定理暗示两个趋势变量之间存在格兰杰因果关系。换句话说，一个变量的创新会驱使或者导致另一个变量的变化路径。该定理的最后一个意义在于，它可以描述趋势变量之间的长期和短期动态平衡。具体做法是运用回归模型来解释时间序列的长期(共同趋势)或短期动态关系(误差纠正机制)(Engle & Granger, 1987)。[19]

我们运用含有两个单位根变量的模型来研究误差纠正模型如何对时间序列之间的长期和短期动态关系进行评估。如果模型中包含此类变量，那么残差将是不稳定的，同时模型参数的推断也是一个非标准分布(不服从 t、f 和 χ^2 分布)(Toda & Philips, 1993; Toda & Yamamoto, 1995)。产生上述问题的主要原因是，在残差中包含没有被模型考虑的趋势。因此，我们需要思考如何恢复这些短期动态关系和长期趋势的信息。为了演示误差纠正模型如何用于$I(1)$变量，请看下列 $I(1)$变量 Y_t 和 Z_t 的单方程误差纠正模型：

$$Y_t = \beta_1 Z_t + \beta_2 Y_{t-1} + \beta_3 Z_{t-1} + \varepsilon_t \qquad [3.31]$$

其中，$\frac{\beta_1 + \beta_3}{1 - \beta_2}$ 是 Z_t 的变化所产生影响的总乘子。这一模型将会被协同整合，也就是根据格兰杰表述定理，在变量 Y_t 和 Z_t 之间建立一个线性组合(Engle & Granger, 1987)。由于方程 3.31 中的残差性具有非稳定性，因此总乘子的系数不能被最小二乘法估计。

对模型中的两个变量求一阶差，我们可以从方程 3.31 推导出模型的误差纠正表述：

$$
\begin{aligned}
Y_t &= Y_{t-1} + \beta_1^* Z_t - \beta_1^* Z_{t-1} + \beta_2^* Y_{t-1} - \beta_2^* Y_{t-2} \\
&\quad + \beta_3^* Z_{t-1} - \beta_3^* Z_{t-2} + u_t \\
(1-L)Y_t &= \beta_1^*(1-L)Z_t + \beta_2^*(1-L)Y_{t-1} \\
&\quad + \beta_3^*(1-L)Z_{t-1} + u_t
\end{aligned} \qquad [3.32]
$$

$$
\Delta Y_t = \beta_1^* \Delta Z_t + \beta_2^* \Delta Y_{t-1} + \beta_3^* \Delta Z_{t-1} + u_t \qquad [3.33]
$$

方程 3.32 存在的问题是，尽管我们运用了一阶差或者稳定数据，但却依然无法恢复方程 3.31 中对总影响乘子的估计信息（未标星号的系数）。为了寻找一种替代方法，我们通过建构下列误差纠正模型来将方程 3.33 变得更加稳定：

$$
\begin{aligned}
(1-L)Y_t &= \beta_1 Z_t + (\beta_2 - 1)Y_{t-1} + \beta_3 Z_{t-1} + v_t \\
(1-L)Y_t &= \beta_1(1-L)Z_t + (\beta_2 - 1)Y_{t-1} + (\beta_1 + \beta_3)Z_{t-1} + d_t \\
\Delta Y_t &= \beta_1 \Delta Z_t + (\beta_2 - 1)\left[Y_{t-1} + \left(\frac{\beta_1 + \beta_3}{\beta_2 - 1}\right)Z_{t-1}\right] + g_t \\
\Delta Y_t &= \beta_1 \Delta Z_t + (\beta_2 - 1)u_{t-1} + g_t
\end{aligned} \qquad [3.34]
$$

方程 3.34 就是模型的误差纠正表达式。对该式进行估计可以同时复原对长期效应乘子和短期效应乘子的估计信息，即使在 $I(1)$ 变量出现的情况下也不会受到影响。与此同时，估计结果也不存在虚假相关的问题，因为残差项 g_t 具有稳定性。

误差纠正模型的估计可以通过两种方法实现。第一种是被称为“双阶段”的方法，即先对长期效应的系数进行估计（方程中的 β_1 和 β_3），然后对误差纠正机制进行估计（β_2）。通过该方法，可以得到一致但可能是无效的系数。第二种方法是运用一个回归方程来完成估计，该回归必须同时包括方程 3.31 中的 Y_{t-1} 和 Z_{t-1}。该方法可以产生既一致又有效的估计系数。一旦估计完成，长期效应就可以通过进一步分析来解决。实践的证据证明，第二种方法对于小样本进行估计的效果更好。

作为向量自回归模型的误差纠正模型

上文所探讨的单方程误差纠正模型实际上是广义二元向量自回归模型的一个简单化。如果我们对 Z_t 明确设定另一个方程，那么我们就可以用向量自回归模型来分析协同整合的可能性。我们设定 $y_t=(Y_t, Z_t)$ 是变量 Y_t 和 Z_t 的 1×2 维向量，那么该模型的向量自回归表达式就是：

$$y_t=\sum_{l=1}^{p} y_{t-l}A_l+u_t$$

用于单方程误差纠正模型的数据变换方法也可以被广泛地用于向量自回归模型。我们先用整个系统减去 y_{t-1} 得到模型：

$$y_t-y_{t-1}=y_{t-1}+\sum_{t=1}^{p} y_{t-l}A_l+u_t$$

$$\Delta y_t=y_{t-1}\Pi+\Delta y_{t-1}\Gamma_1+\cdots+\Delta y_{t-p+1}\Gamma_{p-1}+u_t$$

其中， [3.35]

$$\Pi=-(I_m-A_1-\cdots-A_p)$$

$$\Gamma_i=-(A_{i+1}+\cdots+A_p)(i=1,\cdots, p-1)$$

Δ 代表 y_{t-k} 的一阶差，所以 $\Delta y_{t-k}=y_{t-k}-y_{t-k-1}$ 。这是一个 $p-1$ 阶向量误差纠正模型，它可以让我们复原长期效应（Π）和短期效应 Γ_i 之间的协同整合关系。[20]我们之所以认为这一数据表达形式对于处理协同整合数据来说是最优的，是因为它能够在不需要进一步数据转换的前提下复原对长期效应和短期效应的估计。这些协同整合关系是以变量的线性组合形式出现的，这说明除非协同整合关系的数量被预先选定，否则 Π 将不会是一个满秩矩阵。

对此模型进行估计需要运用一组方程 3.35 的减秩矩阵或者经典的相关分析（Johansen，1995）。该方法通过产生一个对减秩矩阵 $\Pi=\alpha\beta$ 的估计来定义协同整合关系。我们将这一估计结果带入方程 3.35，就可以得到对短期效应 Γ_i 的估计结果。

在设定这一模型时，我们必须回答这样一个问题："模型中存在多少协同整

合向量或关系?”对于含有两个单位根变量的上述模型来说，将会有两个相互独立的随机走动和一个含有一个随机走动的共享趋势模型。这一推断过程需要我们用修改后的非标准分布似然比检验来检验模型中协同整合向量的个数以及向量Π的秩。

每一个进行一阶差处理后的 $p-1$ 阶向量误差纠正模型都可以被写成一个 p 阶的向量自回归模型，具体方法是将前向量误差纠正模型的系数 Γ 和 Π 转换成简化后的向量自回归模型的系数 A_i。这一转换过程可以通过以下步骤完成：

$$A_1 = \Gamma_1 + \Pi + I_m$$
$$A_i = \Gamma_i - \Gamma_{i-1} (i = 2 \cdots p-1)$$
$$A_p = -\Gamma_{p-1}$$

在用向量自回归模型来表示向量误差纠正模型的系数后，前者就包含了后者所蕴含的所有动态关系。对求一阶差后的数据运用向量自回归模型，其反映长期动态关系的效果只能达到向量误差纠正模型减秩估计系数的水平。向量自回归模型的系数无法像误差纠正模型那样区分出模型中的长期和短期效应，因为后者实际上是给前者的系数施加了一系列限制，从而可以被看做一个在时间序列的平衡关系中，对长期效应作出限制的向量自回归模型。

向量自回归模型和向量误差纠正模型的对比

既然两个模型之间的差别并不大，为什么我们在建模过程中要区分两种不同的方式呢？这取决于分析者是只想估计方程系统的动态结构，还是想更进一步地获取协同整合关系的信息。如果我们分析的重点是协同整合关系，那么我们就应该倾向于误差纠正模型。但是，如果只是想评估模型中的短期动态效应和因果关系，那么向量自回归模型则被认为是更有效的。当模型中存在协同整合关系时，运用向量自回归模型检验外生性限定时就必须小心，因为出现的协同整合现象会导致不正确的假设检验(Sims et al.，1990)。

向量误差纠正模型和误差纠正模型对于处理多变量之间存在的相关趋势问题特别有效。这种情况在经济学中十分常见，例如生产力的增长会同时导致

许多变量的增长。但是威廉姆斯的几个观点可能会让我们重新审视这一模型在社会科学中的运用:第一,许多变量可能会出现随机趋势或者确定趋势,但是这些趋势并不一定长期存在;第二,向量误差纠正模型这类建立在单位根基础上的计量经济学方法的经典推论技术值得进一步商榷;第三,误差纠正方法在一些领域的引用范围是很受限制的。

首先,对单位根和误差纠正表述形式的检验依赖于所用的样本。虽然单位根检验第一眼看上去可能无法拒绝虚无假设,但是正如威廉姆斯所认为的,一些变量可能只具有短期趋势(例如总统支持率),是不可能作为单位根或膨胀变量的,因为外生冲击对这类变量的影响会随着时间而消失。因此,尽管样本数据经常呈现出非稳定性,但在具体操作中,如果我们可以利用理论的指导和先验的信息来确定数据具有稳定性,那么就可以将其当做稳定数据来建立模型。

第二,对单位根的检验是一把双刃剑。因为在对单位根进行迪基—福勒增广检验(ADF)和科维亚托夫斯基—菲利普—施密特—申(KPSS)检验时,虚无假设是时间序列中含有单位根(Kwiatkowski, Philip, Schimidt & Shin, 1992)。正如贝叶斯方法所提到的,上述检验对于膨胀或非稳定模型施加了过大的可能性(Sims & Uhlig, 1991),以至于到最后,我们利用经典的检验思想很有可能认为时间序列含有单位根。正如威廉姆斯所指出的:

> 概念上来说,单位根检验的主要困难在于,运用有限时段的样本所判断出来的膨胀或随机趋势可能并不会长期存在。经典的推断方法是从样本数据的特性去推断总体的属性。在时间序列中,样本本来不是随机的,且总体包含着未来和过去两方面的信息。但这一点就足够让我们警惕检验是否能在重复样本中给出精确的估计。经济学往往沉浸于对单位根领域的探讨中,实际上这会使我们对时间序列作出错误的假定。

但必须承认,单位根计量方法对于一些常见的、具有长期趋势性的数据十分有用,例如消费和收入等(Williams, 1993:231)。

威廉姆斯指出的第三点是,我们不能夸大数据中趋势的长期性和普遍性

(William, 1993:232)。误差纠正模型的真正优势在于,它可以被用于简化对具有长期和短期动态过程的多元时间序列模型的解释。向量自回归模型可以被用于描述误差纠正过程和具有更复杂的动态关系的多元时间序列系统。从概括动态特征和描述时间序列之间的关系的角度看,向量自回归模型也可以起部分作用。误差纠正模型不能排除向量自回归模型可以同样出色地完成分析具有长期效应和协同整合性质的数据的可能,因为误差纠正模型仅仅是对长期行为施加了特定限制的向量自回归模型而已。

对向量自回归模型的批评

向量自回归模型对识别预设十分敏感。在无法找到满足结构方程模型设定的预设时,向量自回归模型是一个具有吸引力的替代选择。两者的主要差别在于,后者更多地关注动态性并通过移动平均反应过程来解释变量之间的同期关系。但是向量自回归模型也存在一些问题,其中最大的焦点是对"过度参数化"的批评。因为即便是较小的向量自回归模型,也包含了数量很大的回归参数。一些批评指出,大量的参数导致它们在向量自回归模型中被无效地估计,所以对于统计推论来说是无用的(Pagan, 1987)。这些批评还认为,F 检验和 χ^2 检验可以被用于发展更简化的模型。

这一观点从设定和检验结构方程模型的早期方法演化而来。该方法认为,我们应该设定一个宽泛的模型,然后检验这种零限定是否有意义,其目的在于建构一个更简化的模型。这一方法的逻辑是,简化模型所获得的更多的自由度能够让估计更有效,且在预测、分析冲击反应和政策模拟时,使估计值的置信区间更小。

向量自回归模型的使用者应该意识到,简化原则具有十分显著的作用。实际上,较大的自由度可以让自回归整合移动平均模型变得十分具有解释力,因为该模型的目的就是寻找变量动态关系的最简化的表述。但基于以下两个原因,我们不能使用该方法:其一,我们如何判断零限定的显著性?由于需要依靠先验的但可能是错误的限定来进行检验,这会导致对模型设定过程内在不确定

性的低估，所以对方程的零限定很可能是武断的。尽管简化模型可能更有效，但是最终模型中报告的估计的有效性可能是虚假的，因为它没有考虑到模型的选择和设定过程。其二，向量自回归模型也拒绝用“设定—估计—检验—再设定”的方法来追求模型简化型的逻辑，原因是这个循环会使用二次数据。西姆斯曾经指出，这一做法会产生严重的问题，即让我们对表面上的有效性过度自信，但忽视了这种有效性其实并不真实存在(Sims，1986b、1988)。这种使用数据的方式(第一次将数据用于估计和检验，第二次用于重新设定和再次估计)实际上是两次使用有限的自由度。所以，用模型的结果进行的推论和预测将产生过高的预期。这意味着我们可能对预测和政策分析的结果过于自信，并且可能会把置信区间和 p 值报告得“太好”。

对向量自回归模型的第二个批评是其“缺乏理论”。库利和勒罗伊对该批评的阐述可能是最为清楚的(Cooley & LeRoy，1985)。使用结构方程模型的人之所以摒弃向量自回归模型，主要原因是，自回归模型着眼于模型的简化形式而非模型的整体结构关系。就像我们讨论两种模型之间的关系以及格兰杰因果关系在结构方程模型和在模型的简化形式中是不同的。这一批评在很大程度上是基于将结构方程模型对于外生性的设定作为识别参数时的限定条件。但向量自回归模型的使用者并不接受这一观点，相反，他们所作出的识别假设和要估计的动态关系可能截然不同。这个对向量自回归的批评基于对内生变量和外生变量的划分和限定。尽管这一批评具有理论的吸引力，但是向量自回归模型的使用者却持怀疑态度。

第三个对向量自回归建模和解释过程的主要批评基于库利和勒罗伊认为该模型是非因果的论点。向量自回归模型以及通过移动平均反应过程或创新计算的方法进行结果解释是基于“对因果顺序的条件相关的设定”(Cooley & LeRoy，1985:301)的。这一批评的理由是，在解释内生变量 X 对另一个变量 Y 的影响时，变量 X 相对于 Y 必须具有微弱程度的外生性或者先决性。否则，两个变量之间的动态关系是不能被确定的(从结构方程的角度来看)。

向量自回归模型的使用者也对上述观点作出了一些回应。首先，我们记得模型是基于方程系统的简化形式。因此，冲击反应和创新计算分析所解释的变

量关系是在给定了简化形式或者结构性的误差协方差之后作出的。也就是说，模型的解释所依赖的途径和结构已经被模型中的同期创新明确地包含了。[21]如果对误差协方差矩阵 Σ 进行考利斯基分解时的变量顺序是不正确的，那么对冲击反应和创新计算的条件性预测也将是错误的。

尽管外生性限定可能是向量自回归建模过程的核心，但是该模型的使用者并不在估计前对模型中的外生性限定作出任何假定。该模型运用模型简化的表达形式进行估计，然后对移动平均反应过程和创新计算施加有关变量之间同期创新的因果顺序的假定。这一过程需要模型的使用者对变量之间的关系作出明确的设定。并且，我们在评估移动平均反应和格兰杰因果检验时，必须运用不同的设定。

移动平均反应过程和创新计算是基于跟踪方程系统中变量对外生冲击和创新的反应和变化的。在残差项不存在序列相关的向量自回归模型的简化形式中，外生冲击和创新是不可预测的随机误差，这意味着，它们是不能被系统中的变量预测的，是外生于方程系统的。有关变量中同期关系的问题依然存在，但是通过在不同识别预设条件下检验移动平均反应和创新计算的稳健性，我们可以建立创新之间的相互外生性。

向量自回归模型是单方程的自回归整合移动平均模型、结构方程模型以及误差纠正模型的一个广义形式。实际上，我们可以将后三种模型的单方程情况视为向量自回归的一个特例(详细的讨论请见 Reinsel，1993)。因此，向量自回归模型可以使分析者用更广义的模型来减少对动态关系进行推断的偏误。

向量自回归模型的使用者对既有批评的最后一个回应是，对基本模型的部分修正已经使其能够像结构方程模型一样采取更简化的模型和结构设定，但是这些拓展并不是直接检验对方程的限定，而是运用贝叶斯方法放松对模型的参数的概率限定。向量自回归模型的第二个拓展是结构向量自回归模型，它对模型中的结构性解释和同期因果关系进行了发展。近来的一些工作则是将上述两个新的发展方向进行综合(详见 Leeper，Sims & Zha，1996)。

第 4 章 | 向量自回归分析范例

本章将呈现两个运用向量自回归模型进行多元时间序列分析的完整例子。在每个例子中，我们都会讨论模型是如何被设定、估计和解释的。我们的主要目的是为我们在第 3 章中讨论的模型提供一个更为清晰的范例。这些例子主要是直接向量自回归模型及其推断的通俗易懂的示范。我们将这些例子视为对帕甘的批评(Pagan, 1987)的回应，他认为向量自回归模型的方法论很难被评估和解释。

第一个例子讨论了美国政党参与和公众对政策的偏好之间的动态关系。该模型反应的是斯廷森测量的“公众态度”(Stimson, 1999)和加总后的宏观政党参与之间的关系。这是一个比较基本的例子，可以让我们对模型的基本运用进行探讨。我们承认该模型存在设定不足的现象，但这并不影响我们以其为例子来探讨模型的基本运用。

第二个例子则更为复杂，是威廉姆斯和柯林斯关于政治影响和经济因素对于公司税率的影响(Williams & Collins, 1997)。该研究建立在商业利益团体的政治影响、经济条件和商业税率之间关系的理论预期模型之上。这个例子更高级，其分析也更加复杂。[22]

公众态度和宏观政党参与

美国公众对于政府及其行动的支持与不同政党在参政总体人数中的比例是否有关系？这一问题在有关美国公众意见的文献中十分常见，它包含两个内生概念。一种观点认为，公众期望政府更加主动或者允许政府更加主动，所以

更可能加入当权的政党。因此,公众态度和总体的政党参与率就有了联系。一些研究者已经对这些问题进行过详细的探讨,包括斯廷森、埃里克森、麦奎因。[23]

这一例子中含有两个变量,第一个是公众态度,测量的是公众对政府行为的支持程度。该变量的测量来自盖洛普公司运用动态因子模型对民意调查问题的加总结果(Stimson, 1999)。理论上来说,这一测量基于在跨时段的民意调查中支持美国政府的公众的百分比。具体的测量是将多个理论取值在 0—100 之间的跨时段调查问题进行加总,值越高则代表更加支持政府的行动,反之亦然。第二个变量,宏观政党参与,是根据盖洛普民意调查的数据计算的民主党的党员数量在两党的党员中所占的比例。我们的例子用的是 1958 年第四季度到 1996 年第四季度之间的季度数据。

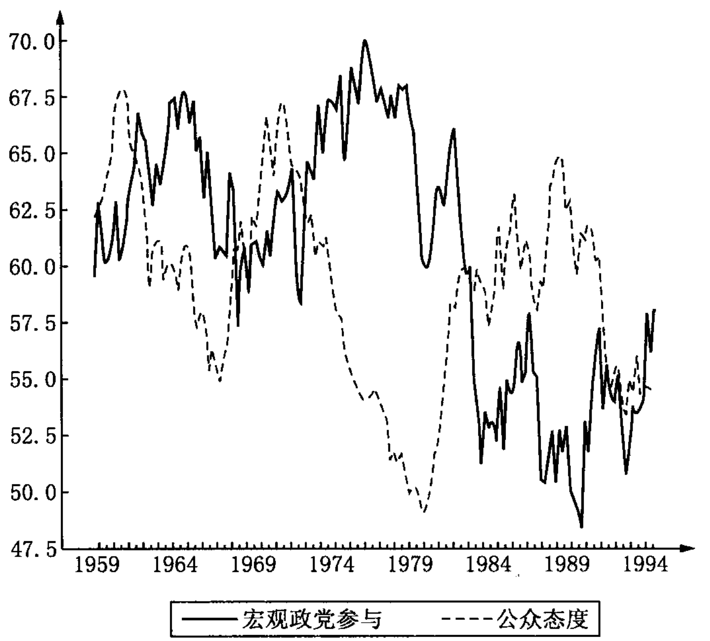


图 4.1　宏观政党参与率和公众态度的季度变化(1958 年 4 月—1996 年 4 月)

图 4.1 呈现了两个变量的时间序列。可以看出,两个序列都具有局部趋势,即在特定的较短时间段内,呈现出比较一致的升高和降低趋势。因此,除了检验向量自回归模型的滞后时长之外,我们还必须对数据是否存在单位根或者趋

势变量进行检验。我们将执政党、众议院和参议院中各党的党员数量作为控制变量。如果这些变量中民主党占优势，则标记为 1；如果共和党占优势，则标记为 0。对模型的设定和分析遵循以下步骤：(1)对单位根进行检验；(2)对滞后时长进行检验；(3)设定和估计向量自回归模型；(4)对格兰杰因果进行检验；(5)对预测误差方差的创新计算进行分解；(6)对冲击反应作出分析和解释。

我们对每个步骤都会进行细致的探讨，然后用现实的例子来讲解如何解释向量自回归的估计结果。

单位根的检验

我们首先检验两个变量的单位根现象。这一检验之所以重要，是因为如果变量含有单位根，那么格兰杰因果检验中的误差纠正机制就有可能是错的。对单位根的检验是通过 ADF 检验完成的(细节请见 Hamilton, 1994)。表 4.1 报告了滞后项从 0 增加到 8 的 ADF 检验结果。

ADF 检验的虚无假设是模型中有一个单位根，该单位根包含一个随机走动和一个常数，检验的临界值大约是－2.88。我们可以从表 4.1 中看出宏观政党参与变量中存在一个随机走动，统计量小于临界值，因此我们不能拒绝虚无假设(Box-Steffensmeier & Smith, 1996)。但对于变量是否含有趋势性的问题，检验结果无法拒绝原假设且对趋势的估计是统计上不显著的。对于公众态度变量来说，决定趋势是不显著的，而 ADF 检验结果也指出，单位根的虚无假设对于含有更多滞后项的模型来说更容易拒绝。

表 4.1 广义迪基—福勒检验

滞后	宏观政党参与		公众态度	
	无趋势	有趋势	无趋势	有趋势
0	－2.04	－2.97	－1.75	－1.92
1	－1.65	－2.33	－1.87	－2.03
2	－1.55	－2.24	－1.93	－2.09
3	－1.6	－2.46	－2.08	－2.19
4	－1.79	－2.73	－2.49	－2.59
5	－1.54	－2.45	－2.67	－2.75

续表

滞　后	宏观政党参与		公众态度	
	无趋势	有趋势	无趋势	有趋势
6	−1.48	−2.22	−2.57	−2.55
7	−1.3	−2.25	−2.31	−2.28
8	1.08	−2.1	−2.82	−2.79

注：在5%的显著性水平下，对无趋势性检验的临界值大约是−2.88，对有趋势检验的临界值大约是−3.50。

宏观政党参与变量可能会出现单位根，但是该变量存在上下限，因为这一变量是盖洛普民意调查中，民主党人数占总参政人数（民主党和共和党）的百分比。所以该变量肯定在0—100这个范围之内，并且在较长时间范围内是具有稳定性的。这一事实可能与单位根检验的结果不符，因为检验结果暗示，宏观政党参与的可能性在较短的一段时间内是不稳定的，可能存在局部趋势或者常数。但是从长时段的角度来看，该变量具有稳定性，因为当民主党的参与率达到一个很高的百分比时，它一定会下降。因此，正如威廉姆斯所建议的，上述经验和理论让我们可以将这一变量视为稳定的。[24]

设定滞后时长

对向量自回归模型滞后时长的设定一般是两个部分的组合，一个是拟合统计量，另一个是对滞后时长的正式检验统计量。大家可能记得，任何有限阶滞后时长只能近似于一个可能的无限阶滞后时长。因此在实际分析中，我们可以对滞后时长检验以及拟合检验采取反对态度。

表4.2　宏观政党参与率与公众态度的向量自回归模型的赤池信息准则和贝叶斯信息准则

滞后	AIC	BIC	滞后	AIC	BIC
1	238.24	255.58	7	259.53	346.24
2	240.86	269.77	8	259.11	357.38
3	246.69	287.15	9	255.47	365.3
4	253.75	305.78	10	256.07	377.47
5	251.26	314.85	11	260.29	393.24
6	252.95	328.1	12	263.00	407.52

表 4.2 呈现了从含有 1 个到 12 个滞后项的向量自回归模型的赤池信息准则和贝叶斯信息准则的值。这些数字可以被解释为描述模型的似然函数在对增多的滞后项进行调整之后，似然函数增长了多少。这两个统计量的取值越小，模型的拟合度就越高。基于表 4.2，我们选择滞后长度为 $p=1$。我们也可以选择 2 个滞后项，但是检验结果表明，两者之间不存在什么差异(相对于 2 个滞后项和 3 个滞后项的差别来说)。虽然我们担心残差项中的序列相关问题，但是加入更多的滞后项会提高对估计效度的限定要求。基于这些原因，我们只采用 1 个滞后项。请注意，随着加入更多的滞后项，似然函数值并没有发生什么变化。我们在第 3 章中讨论过，每一个新加入的滞后项给赤池信息准则带来的惩罚因子是 $2m^2=8$，而对贝叶斯信息准则是 $\log(T)m^2=\log(124)2^2=8.37$。如果加入 1 个滞后项带给两个信息准则的变化小于这个值，那么它就无法给模型带来足够的改善，也就没有加入的必要了。[25]

另一个评估滞后时长的方法是运用假设检验。表 4.3 得出了同样的结论。χ^2 检验比较的是多加 1 个滞后项的模型和原有模型。证据表明，我们不能拒绝 2 个滞后项的模型而倾向于 1 个滞后项的模型(P 值等于 0.27)，但是检验结果却让我们足以拒绝含有 3 个滞后项的模型(P 值等于 0.72)。基于这些结果，随后的向量自回归模型只包含 1 个滞后项。[26]

表 4.3 滞后时长的似然比检验

非限定滞后时长	限定滞后时长	卡方	P 值
12	11	4.29	0.37
11	10	3.13	0.54
10	9	6.23	0.18
9	8	9.98	0.04
8	7	7.35	0.12
7	6	1.26	0.87
6	5	5.7	0.22
5	4	9.62	0.05
4	3	0.88	0.93
3	2	2.06	0.72
2	1	5.18	0.27

注：χ^2 检验的自由度是 4。

向量自回归模型的估计

模型中不仅包含了1个滞后项,还有3个控制变量。控制变量是总统职位、参议员和众议院中哪个政党更占优势的二分变量。如果是共和党人出任总统,或者在参众两院占据更多席位,那么变量就会被记为“0”,相反,如果是民主党更占优势,则标记为“1”。

表 4.4　含有1个滞后项的公众态度与宏观政党参与向量自回归估计

变　量	因变量	
	公众态度	宏观政党参与率
公众态度$_{t-1}$	0.939(0.029)	0.013(0.042)
宏观政党参与率$_{t-1}$	−0.019(0.022)	0.954(0.031)
总统$_t$	−0.424(0.282)	−0.041(0.409)
参议院$_t$	0.077(0.614)	−1.069(0.889)
众议院$_t$	−0.187(0.348)	0.410(0.504)
常数	4.900(2.233)	2.743(3.232)
R^2	0.919	0.897
标准误	1.290	1.867
Durbin-Watson	1.952	2.381

注:括号内为标准误。

模型中每个方程含有6个系数,分别是总统的政党身份、控制参议院的政党、控制众议院的政党以及这些变量在前一个时段的值。表4.4报告了向量自回归模型系数以及标准误。前四行的自回归系数描述了系统的动态性。这些系数并不具有单独的意义,因为它们表示的是系统的行为,所有的系数都是在描述多个变量的动态关系。关于这些参数的任何推断都包含第一个和第二个滞后项的4个系数。由于系统的动态性,向量自回归模型的系数是一个需要评价其稳定性的矩阵。

我们先解释控制变量的意义,即控制总统职位和参众两院的政党。估计系数描述了控制总统职位和国会席位的政党对于每个内生变量平衡的方向。只

有公众态度方程中的总统职位变量($t=1.5$)和宏观政党参与方程中的参议院控制变量($t=-1.2$)接近显著。我们可以看到,民主党人出任总统会降低公众对政府行为的支持,同时,民主党控制参议院也会抑制其总体的党员数量。这些对宏观政党参与和公众态度的估计将使动态关系变得更为复杂。尽管系数的符号可以被解释,但是外生变量大小的变化取决于系统的动态性,所以我们需要预测误差分解和冲击分析等替代方法来解释模型并进行推断。

格兰杰因果检验

在表 4.4 中的估计值以及残差的 2×2 协方差矩阵可以被用于检验系统中是否存在格兰杰因果。在第 3 章的讨论中,我们提及格兰杰因果可以通过 F 检验来评估。我们在这里依然只包含一个滞后项。

表 4.5　对宏观政党参与率和公众态度在向量自回归模型中的格兰杰因果检验

假设的外生变量	限定系数组	F 统计量	P 值
公众态度	宏观政党参与率	0.74	0.39
宏观政党参与率	公众态度	0.09	0.76

表 4.5 是对向量自回归模型进行格兰杰因果检验的结果。检验是为了确定滞后的公众态度变量在宏观政党参与的方程中是否为 0,以及滞后的宏观政党参与变量在公众态度的方程中是否为 0。

对于两个外生性检验来说,我们无法拒绝原假设。因此证据表明,公众态度并不依赖于宏观政党参与的过去值,反之亦然。这一结果将会引发一个问题,因为如果两个变量之间不存在动态性影响,那么它们之间的关系就可能是同期性的相关,但是在单位根出现之时,格兰杰因果检验是有偏的(Hamilton, 1994: 554)。

我们可以通过表 4.4 中对向量自回归模型的估计来看这一问题。我们发现,自回归系数矩阵接近一个单位矩阵。这是寻找单位根变量的模拟矩阵。我们需要将可能存在单位根的有偏的格兰杰因果检验结果和含有不同滞后时长模型的格兰杰因果检验结果相比较,这是因为格兰杰因果检验对这两个因素都

很敏感。其结果显示，滞后时长从1增长到8并没有改变检验结果，所以这和我们在表4.5中得出的结果是相同的。

预测误差方差分解

预测误差分解旨在解释模型的拟合是如何随着内生变量的向量的实际取值而改变的。具体做法是，运用向量移动平均式来表示向量自回归模型，然后计算跨时段的预测离差。预测误差的方差随后被分解到每个内生变量上，即每个自变量解释了多少方差。如果这些变量彼此外生，我们的预期是一个变量的创新并不会解释其他变量的方差。但如果它们彼此同期相关，那么我们就认为一个变量的方差可以通过将一个滞后项作为同期创新，并通过方程系统的其他滞后项来解释剩余变量。

对公众态度和宏观政党参与的预测误差分解结果显示在表4.6中。前两列是外生冲击或者创新导致的对公众态度变量的预测方差百分比。在第3章的讨论中，我们知道，这一分解过程基于模型残差的协方差矩阵。因为公众态度变量处在变量序列中的第一位，所以分解在开始阶段假设，预测中所有的方差都来自公众态度。随着预测范围的增加，更多的方差被归于系统中其他变量的创新或者它们之间的相关关系。在这个例子中，10个季度也就是2.5年之后，公众态度的预测方差中有1.75%可以被归于宏观政党参与的创新。这一趋势在这个时段之后趋于稳定，所以在16个季度之后，公众态度的预测误差中有3.95%来自宏观政党参与的创新。

表4.6 对公众态度和宏观政党参与的向量自回归模型的预测误差平方分解

k	公众态度的创新预测误差率(%)		宏观政党参与的创新预测误差率(%)	
	公众态度	宏观政党参与率	公众态度	宏观政党参与率
1	100.000	0.000	2.755	97.245
2	99.963	0.037	2.904	97.096
3	99.878	0.122	3.052	96.948
4	99.750	0.250	3.197	96.803
5	99.582	0.418	3.340	96.660

续表

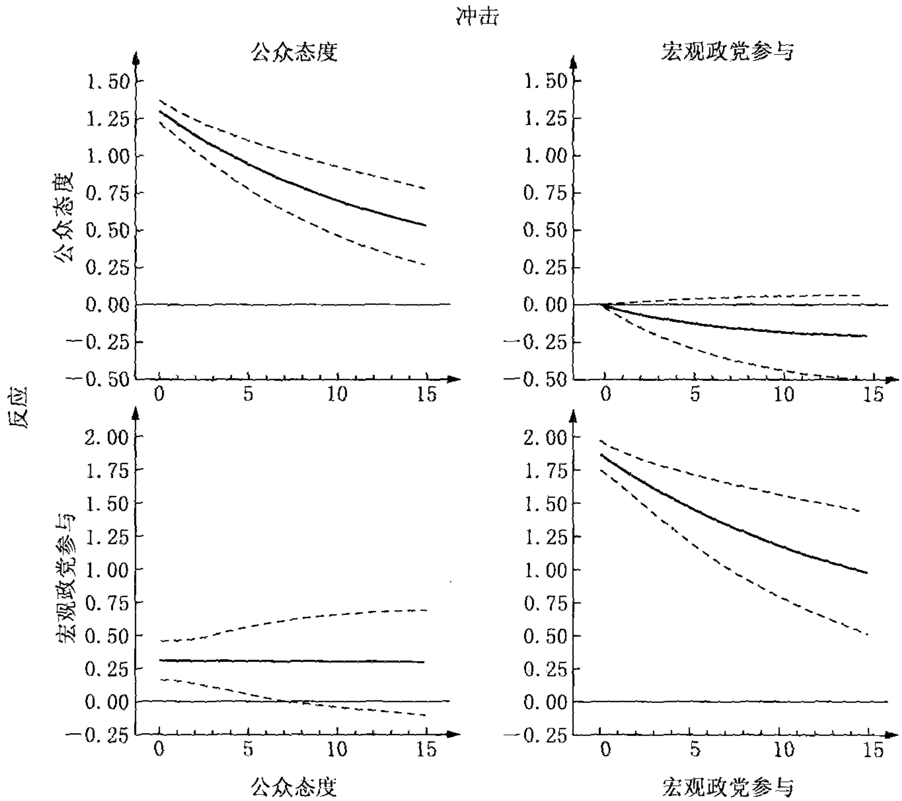
k	公众态度的创新预测误差率(%)		宏观政党参与的创新预测误差率(%)	
	公众态度	宏观政党参与率	公众态度	宏观政党参与率
6	99.376	0.624	3.480	96.520
7	99.136	0.864	3.617	96.383
8	98.866	1.134	3.751	96.249
9	98.570	1.430	3.881	96.119
10	98.251	1.749	4.008	95.992
11	97.913	2.087	4.130	95.870
12	97.559	2.441	4.249	95.571
13	97.192	2.808	4.364	95.636
14	96.817	3.183	4.475	95.525
15	96.435	3.565	4.581	95.419
16	96.049	3.951	4.684	95.316

表4.6最右边两列是对宏观政党参与的预测误差分解。在16个季度之后，大约有5%的预测误差可以归于公众态度的创新。[27]

从上述预测分解分析中，我们可以得出结论：宏观政党参与中无法预期的变化对公众态度创新的影响很小，不会大于方差的4%。反之，公众态度创新对宏观政党参与的预测方差也很小。实际上，我们看到，两个变量之间的相互影响是慢速且微弱的。这一结果符合我们的预期，即偏好导致政策选择。我们在考利斯基误差协方差分解中重新排列变量的顺序，依然得到了同样的结果。

冲击反应分析

冲击反应分析能让我们在向量移动平均表达式中，对向量自回归模型的动态关系进行分析。从本质上来说，这一方法能够让我们追踪外生冲击所造成的内生变量的变化是如何随时间而变化的。向量移动平均过程的设定以及随后的冲击反应分析涉及一个问题，即同期关系的顺序。在这里，同期相关及其顺序是通过对残差协方差矩阵的考利斯基分解来完成的。[28]



注：公众态度在变量顺序中处于第一位。误差带的宽度约为68%的置信区间或者1个标准差。

图4.2　对宏观政党参与和公众态度的冲击反应分析

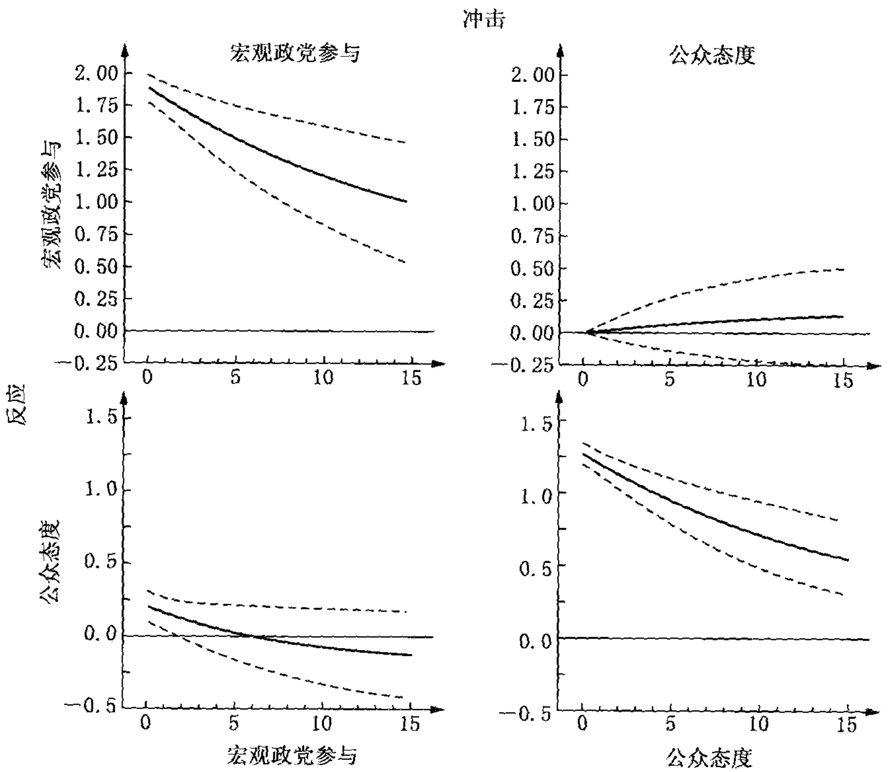
图4.2呈现的冲击反应的结果是将公众态度放在变量序列中的第一个时，所做的对残差的同期协方差分解。这意味着，外生冲击对宏观政党参与在最初阶段的影响为0。所以，在第一行中列出的是公众态度对两个外生冲击的反应。我们可以看出，1.25个冲击单位(公众态度方程残差的一个标准差)在16个季度之后大约减少了一半。右上角的图显示了公众态度对宏观政党参与的冲击的反应，该反应大约为1个标准差(或者说68%的置信区间)。在第3章中我们提到，图形是对冲击反应集中趋势的一种更好的概括方法。公众态度对于宏观政党参与中的创新的反应在16个季度后变得很弱，并且我们有68%的置信区间包含0。

第二行的两个图是当两个变量在序列中具有同样的位置时，宏观政党参与对于外生冲击的反应。我们可以看出，对公众态度的1.25个冲击单位导致了在

后续的 6 个季度中，宏观政党参与以 0.25 的稳定速度增长。这个效果在 6 个季度后变得不显著，置信区间也包含 0。最后，右下角的图显示了对宏观政党参与的冲击。由于这个变量中存在很强的类似单位根的现象，所以其冲击持续了很多季度且减少得很慢。

这些结果告诉了我们有关宏观政党参与和公众态度之间动态反应的两个事实。首先，公众态度对于宏观政党参与的反应十分弱。相反，宏观政党对公众态度却存在持续的反应。第二，上述结果主要是由外生冲击中较弱的同期相关导致的。用于计算残差协方差矩阵的同期残差之间的相关系数为 0.17。

但是，我们对于冲击反应中变量序列的顺序设定并没有参照政治学的理论。所以，我们应该继续分析当宏观政党参与处在变量序列的首位时的分解情况。图 4.3 就是这一分解的结果。



注：公众态度在变量顺序中处于第一位。误差带的宽度约为 68%的置信区间或者 1 个标准差。

图 4.3 对宏观政党参与和公众态度的冲击反应分析

图4.3是宏观政党参与和公众态度对其自身冲击的反应。沿着从左到右对角线的两张图与图4.2中同样位置的两张图完全一样,发生变化的只是从右到左对角线的两张图。在第二阶段的分解中,政党参与对外生冲击的反应很小且迅速失去了统计上的显著性(图4.3中的左下图)。而公众态度对冲击的反应尽管在统计上不显著,但却是增加的(在68%的误差范围内)。

总的来说,变量的顺序对于冲击反应的影响并不大。如果我们认为公众态度的冲击先于政党参与,那么前者1个标准差的变化就会导致后者0.25点的变化。相反,如果变量的先后关系掉转,那么反应就会小一些。在两种情况下,证据证明变量是同期相关的,且一个变量对另一个变量冲击的反应是微弱且缓慢的。

有效公司税率

我们的第二个例子是较为复杂的向量自回归分析。该例子中含有更多的动态关系、单位根变量和更具有理论性的模型设定以及因果推断。在本例中,我们部分地复制威廉姆斯和柯林斯的分析(Williams & Collins, 1997),他们运用向量自回归模型分析了美国公司税率的决定因素。一些人认为,税率是对政治压力的反应。共和党控制国会和白宫时将会通过公司政治行动委员会来对国会议员施压,从而制定较低的商业税率。但是威廉姆斯和柯林斯认为,这一政治经济学模型预测的最优税收政策是外生于政治变量的,但是它们却受到经济变量的影响。

威廉姆斯和柯林斯提出的税收模型认为,一些政治变量,例如公司政治行动委员会的贡献以及该商业性的委员会相对于其他政治行动委员会的规模,是外生于有效公司税率的。该理论认为,商业活动对投资和经济的影响是税收政策的根源。因此,税收政策应该是被根本的经济因素决定的,而非商业团体的正式压力。这个模型针对公司税率、政治行动委员会中公司的数量以及经济状况提出了3个命题:

命题 1：有效公司税率外生于商业团体的政治能力。相反，有效公司税率的增长会促进商业政治团体有组织的行动。

命题 2：投资对有效公司税率很敏感，对有效公司税率的外生冲击会降低投资水平。

命题 3：有效公司税率外生于经济条件，包括实际投资总额和实际收入。

我们运用向量自回归模型来检验 4 个变量之间的关系。命题 1 是预测格兰杰因果的方向。有效公司税率被假定外生于商业团体过去的政治能力。命题 2 陈述了投资对税率外生冲击的反应，即投资会逐渐降低。这意味着冲击反应和预测误差方差分解可以描述这一过程。最后一个命题是经济变量的过去值不会影响当下的税率，我们可以用格兰杰因果检验来验证该命题。

数据

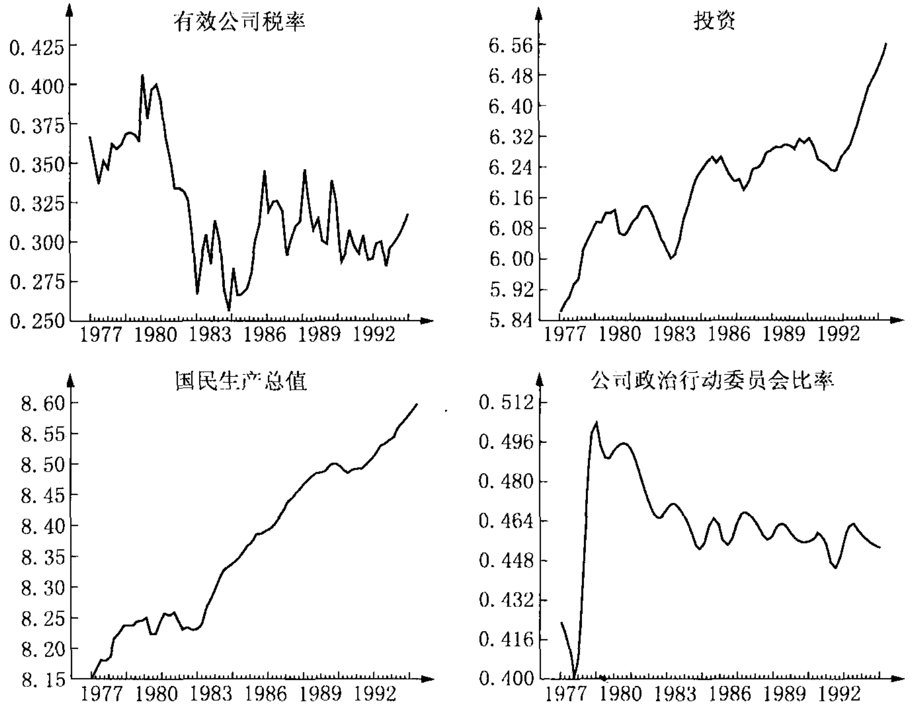


图 4.4　有效公司税率数据

威廉姆斯和柯林斯运用的数据是从 1977 年第一季度到 1994 年第四季度。这是因为政治行动委员会的数据在 1977 年初的财政改革运动之后才公布。该分析涉及 4 个变量,它们是有效公司税率、国民生产总值的自然对数、投资的对数以及商业性质的政治行动委员会在所有委员会中的比率。图 4.4 是根据所用数据绘制而成的。

对于这些数据,我们需要考虑的是设定向量自回归模型时的可能趋势。图 4.4 告诉我们,数据可能含有单位根,尤其是在国民生产总值这个变量中。接下来,我们将会讨论对单位根的检验和滞后时长的设定。

单位根的趋势

诸如国民生产总值一类的经济变量一般都含有单位根。所以问题的关键是,数据中的单位根是否意味着变量之间的协同整合关系。表 4.7 是对模型中每个时间序列的 ADF 检验。

表 4.7　有效公司税率的单位根检验

滞后项个数	有效公司税率	国民生产总值	投　　资	政治行动委员会比率
1	−2.92145	−0.56335	−0.42694	−5.11362
2	−2.78095	−0.71887	−0.40087	−3.60019
3	−2.55360	−0.95425	−0.48052	−5.03627
4	−2.06522	−1.09437	−0.61641	−3.99344
5	−1.92103	−1.20439	−0.59058	−1.72695
6	−1.93909	−1.06926	−0.46436	−2.18983
7	−1.86593	−1.03568	−0.61092	−3.22650
8	−1.65129	−0.83706	−0.54889	−3.13094

注:表格中的数据是 ADF 检验结果。

检验的临界值是−2.88,所以在有效公司税率和政治行动委员会比率中不存在单位根。相反,国民生产总值这个变量一定含有单位根。这意味着对统计检验的解释必须十分小心,因为数据中可能存在协同整合关系。[29]

尽管存在出现协同整合关系的可能性和相应证据,但这并不足以让我们选

择误差纠正模型或者向量误差纠正模型。这是出于两个原因:首先,我们预期如果在国民生产总值和投资变量中含有协同整合关系,那么系统将会是稳定的,并且统计检验将会具有渐进有效性。第二,协同整合关系并不是模型和命题关注的重点。短期的动态关系和外生性分析才是我们关注的中心,而这两个方面在出现单位根的情况下依然是稳健的。

设定滞后时长

考虑到季度数据的特征,常用的滞后时长是从 6 开始,到能够解释数据季节性的滞后时段结束。在下文的分析中,我们以 6 个滞后项进行分析。为了证明这一做法的正确性,我们在表 4.8 中列出了赤池信息准则和贝叶斯信息准则,滞后项的个数从 1 到 12。

表 4.8 有效公司税率向量自回归模型的赤池信息准则和贝叶斯信息准则的滞后时长检验

滞后项个数	AIC	BIC	滞后项个数	AIC	BIC
1	−2220.07	−2178.19	7	−2345.59	−2102.65
2	−2298.53	−2223.13	8	−2403.67	−2127.22
3	−2315.44	−2206.53	9	−2442.30	−2132.34
4	−2322.42	−2180.00	10	−2466.42	−2122.95
5	−2315.78	−2139.85	11	−2500.43	−2123.45
6	−2349.06	−2139.63	12	−2548.22	−2137.73

正如我们在表 4.8 中看到的,从第一个到第四个滞后项,赤池信息准则一直在减小。该数字从第五个滞后项开始增加,到第六个滞后项又开始降低。这是因为第一个"低"值出现在第四个和第六个滞后项之间,以这个证据可以确定适当的滞后时长。同样,贝叶斯信息准则从低到高的排列顺序也是 2、3、4、1 和 6 个滞后项。因为第四个和第六个滞后值之间的差别与最低取值的信息准则取值区别不大,我们选用 6 个滞后项来确保残差项中没有序列相关。

表 4.9　滞后时长的似然比检验

非限定滞后时长	限定滞后时长	卡方	P值
12	11	14.63	0.55
11	10	16.5	0.42
10	9	17.77	0.34
9	8	27.07	0.04
8	7	40.53	<0.01
7	6	14.74	0.54
6	5	38.08	<0.01
5	4	16.48	0.42
4	3	27.94	0.03
3	2	38.31	<0.01
2	1	93.89	<0.01

注：统计量服从经过小样本纠正的、自由度为 16 的 χ^2 检验。

另一种方法是看滞后时长的似然比检验结果。表 4.9 呈现了连续滞后时长的似然比检验结果。但需要注意的是，由于单位根的出现，使得经典的假设检验方法对一些变量失效，所以我们并不会看到稳健的检验结果。表格先对第一个和第二个滞后项进行比较，然后再比较第二个和第三个，依次类推。可以看出，该结果与赤池信息准则和贝叶斯信息准则的结果相同，所以 4 个和 6 个滞后项我们都会使用，一是与我们分析的季度数据相符，二是可以用 6 个滞后项来描述残差项中的季节效应。[30]

格兰杰因果检验

对于变量之间关系的探讨，需要求助于格兰杰因果关系。二元格兰杰因果关系是通过一系列二元向量自回归模型来对主要变量进行评估。[31]表 4.10 是外生性检验结果。

该表呈现了对税率和委员会比率两个变量一阶差的外生性检验。之所以这样做，是因为税率变量中可能存在单位根。表格的第一列是假设为外生的变量，第二列是在格兰杰非因果关系的虚无假设下被限定为 0 的变量的系数。后两列是 F 统计量和 P 值，其结果为非显著性，所以我们不能拒绝非因果的虚无假设。

既然 F 检验说明不能拒绝非因果假设，也就是说，税率是外生于其他变量且不会由其他变量而导致的。在 0.05 的显著性水平下，税率是外生于公司政治行动委员会比率的。唯一让人产生质疑的是商业性政治行动委员会的总数(对 4 个滞后项的模型来说，$P = 0.18$)。由于这一变量存在趋势性，可能会导致对虚无假设的统计检验有偏，所以我们应该运用不同的滞后项来对模型进行检验。运用 6 个滞后项的模型支持了含有 4 个滞后项的模型。

表 4.10 的结果支持税率格兰杰导致委员会比率这一变量，因为非因果假设被拒绝了(在含有 4 个滞后项的模型中，$P = 0.01$；在含有 6 个滞后项的模型中，$P = 0.12$)。

表 4.10 滞后时长的似然比检验

假设外生变量	限定系数组	F 统计量	P 值
滞后 4 个季度的结果			
有效公司税率	政治行动委员会中公司的个数	1.62	0.18
政治行动委员会中公司的个数	有效公司税率	0.31	0.87
有效公司税率	商业性政治行动委员会占所有委员会的比例	0.40	0.81
商业性政治行动委员会占所有委员会的比例	有效公司税率	9.51	<0.01
求一阶差后的税率	求一阶差后的政治行动委员会中公司的个数	0.42	0.79
求一阶差后的政治行动委员会中公司的个数	求一阶差后的税率	0.43	0.78
求一阶差后的税率	求一阶差后的商业性政治行动委员会占所有委员会的比例	0.36	0.84
求一阶差后的商业性政治行动委员会占所有委员会的比例	求一阶差后的税率	0.35	0.84
滞后 6 个季度的结果			
有效公司税率	政治行动委员会中公司的个数	0.97	0.45
政治行动委员会中公司的个数	有效公司税率	0.50	0.80
有效公司税率	商业性政治行动委员会占所有委员会的比例	1.86	0.11

续表

假设外生变量	限定系数组	F统计量	P值
商业性政治行动委员会占所有委员会的比例	有效公司税率	1.79	0.12
求一阶差后的税率	求一阶差后的政治行动委员会中公司的个数	0.83	0.55
求一阶差后的政治行动委员会中公司的个数	求一阶差后的税率	0.21	0.97
求一阶差后的税率	求一阶差后的商业性政治行动委员会占所有委员会的比例	0.83	0.55
求一阶差后的商业性政治行动委员会占所有委员会的比例	求一阶差后的税率	1.44	0.21

注：结果基于对二元向量自回归的估计。

因此，税率的改变会导致商业行动的变化，这直接支持了命题1。但税率并不会格兰杰导致商业性政治行动委员会的总数，这一结论和委员会数量随时间而稳定增长的趋势相一致。所以，这种趋势变量会导致F检验的失效。但是，运用公司政治行动委员会在所有委员会中的比率测量该变量后，结果和命题1是一致的。

表4.11 对有效公司税率、实际投资和实际收入的外生性检验（1953—1994年、1960—1994年、1977—1994年）

假设外生的变量	限定系数组	4个滞后项		6个滞后项		8个滞后项	
		F	P	F	P	F	P
1953—1994年							
有效公司税率	实际投资	1.27	0.28	1.05	0.40	1.15	0.33
有效公司税率	实际收入	1.77	0.14	1.44	0.20	1.82	0.08
求一阶差后的有效公司税率	求一阶差后的实际投资	1.02	0.40	1.12	0.35	1.02	0.42
求一阶差后的有效公司税率	求一阶差后的实际收入	0.85	0.50	1.59	0.15	1.98	0.05
1960—1994年							
有效公司税率	实际投资	0.72	0.58	0.82	0.56	0.91	0.51
有效公司税率	实际收入	1.13	0.35	1.24	0.29	1.54	0.15

续表

假设外生的变量	限定系数组	4 个滞后项		6 个滞后项		8 个滞后项	
		F	P	F	P	F	P
求一阶差后的有效公司税率	求一阶差后的实际投资	0.80	0.53	0.96	0.45	0.87	0.54
求一阶差后的有效公司税率	求一阶差后的实际收入	0.76	0.55	1.49	0.19	1.65	0.12
1977—1994 年							
有效公司税率	实际投资	0.51	0.73	0.36	0.90	0.76	0.64
有效公司税率	实际收入	0.10	0.99	0.46	0.84	1.06	0.41
求一阶差后的有效公司税率	求一阶差后的实际投资	0.68	0.61	0.52	0.79	0.56	0.80
求一阶差后的有效公司税率	求一阶差后的实际收入	0.28	0.89	0.95	0.47	1.35	0.24

表 4.11 是对税率和对数形式的投资以及收入之间的外生性检验，表述的方法和表 4.10 一样。这些对各种滞后时长的检验得出的结论是，不能拒绝虚无假设，换句话说，向量自回归模型中收入和投资的系数为 0。随后，我们用层级数据和一阶差数据来检验数据中是否存在单位根，结果也显示税率是外生于投资和收入的。这两个结果支持了命题 3，因为税率是外生于经济状态的。[32]

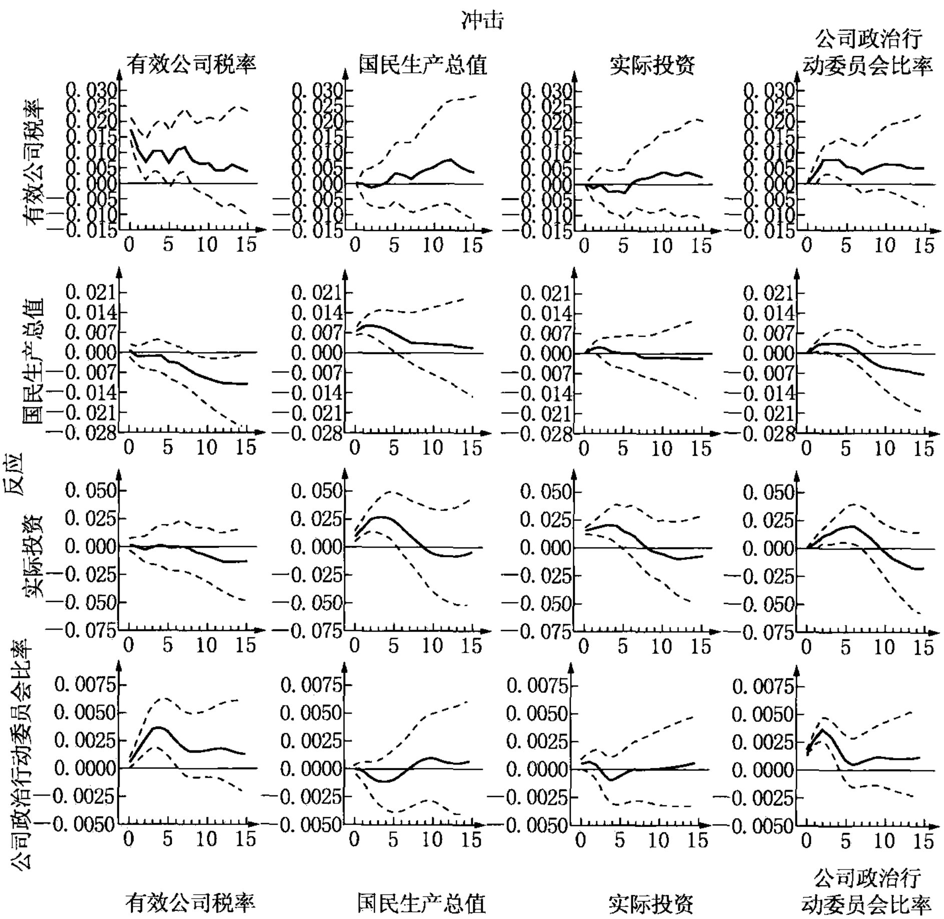
表 4.11 的结果不能用于推断命题 2 是否成立，因为税率的变化会导致投资的变化。这是因为在双变量向量自回归模型中，对外生性的检验过程中可能会出现虚假的证据支持因果关系。命题 2 陈述的是因果关系不是外生的，因此这取决于系统内变量之间的动态关系。在下一章中，我们将分析这些动态关系并运用其他方法来评估命题 2。

冲击反应分析

我们已经在有可能出现单位根的情况下，建立了变量之间明确的外生或因果关系。为了确定方程系统内部的动态关系，我们运用包含 4 个变量的向量自回归模型。我们将包含 6 个滞后项来保证残差项中不存在序列相关。

这一个向量自回归模型是非限定性的，换句话说，我们没有施加任何有关

变量之间结构和外生性的假设，甚至上文已经证明的结论，我们也并不将其作为参考。这对于向量自回归模型来说十分关键，因为该模型的基础是通过对变量之间关系的外生性检验来对模型作出各种限定，但是这些限定条件对于结构模型来说是无效的。为了发现模型中包含哪些动态关系，我们只能分析模型对外生冲击的反应，而这只能通过非限定性的向量自回归模型来得到。我们将估计后的向量自回归模型进行转置，然后用图 4.5 来表示其移动平均反应过程，这是从威廉姆斯和柯林斯（Williams & Collins，1997）的研究中复制过来的。但是图 4.5 的误差范围比威廉姆斯和柯林斯报告的小一些，因为我们运用的是西姆



注：移动平均反映的时间范围是 16 个季度。

图 4.5 4 个变量在 90%误差范围的移动平均分析(1977—1994 年)

斯和扎讨论的似然基础误差范围(Sims & Zha, 1998),而威廉姆斯和柯林斯则基于近似正态方法。因此,图 4.5 的误差范围描述的是建立在对冲击反应的蒙特卡洛模拟后的 90%的分位数上。可以看出,这一结果略微偏大,不仅反应在参数的不确定性上,还包括总体的似然函数的不确定性以及移动平均反应过程的形状和偏态。

图中的行代表的是包含列变量的方程受到 1 个标准差的外生冲击后的反应。这一冲击随后通过转置后的向量自回归方程系统反馈,进而产生对冲击的反应。对角线上的图显示了变量对其自身冲击的反应,因此反对角线上的图则是变量对彼此冲击的反应。这些冲击进入系统的顺序取决于变量在进行冲击反应分解时的顺序。因此在本例中,对税率的冲击率先进入方程系统,然后是国民生产总值、实际投资,最后是公司政治行动委员会比率。

移动平均反应图支持命题 1,因为对税率的冲击导致了商业性政治行动团体比率的升高。命题 2 陈述了投资如何对税率的外生冲击产生反应,从图中所得到的支持很弱,虽然总体的反应是负的,但 90%的单侧置信区间都包含 0。最后,移动平均反应过程图支持命题 3,因为 90%的置信区间都包含 0。这就意味着,税率不会对收入和投资的外生冲击作出明显的反应。请注意,对商业性政治行动委员会的比率的冲击会导致税率的增加,尽管在 8 个季度后变得不显著。威廉姆斯和柯林斯提供的解释是,税率对商业委员会比率的冲击的正向反应是符合预期的(Williams & Collins, 1997:230)。我们看到,经济变量会在委员会比率受到冲击后增长,这会导致公司和商业团体对税率增长产生强烈的预期。这一解释反映的是,模型中没有包含的动态关系,因为我们没有看到国民生产总值、收入等经济变量对委员会比率受到外生冲击后作出显著的反应。

移动平均反应过程为复杂的动态系统提供了一个清晰的图形概括,从而很好地解释了分析中 4 个变量之间的关系。这一方法的主要益处在于,它允许研究者分析方程系统中的动态关系是如何彼此相连的。第二个益处在于,结构方程中运用的那些有关变量之间结构关系的设定,可以通过该模型的外生性检验而得出。我们可以看到,移动平均反应的外生性假设如何受到数据的支持或者被数据拒绝。结合格兰杰因果分析,这些方法能够为我们在处理变量的动态关

系设定和识别问题时提供简洁清晰的途径。

预测误差方差分解

我们用分解预测误差方差的分析来对向量自回归模型及其结果进行最后的解释。[33]

表 4.12 向量自回归模型误差方差分解

预测误差		创新			
	k	有效公司税率	投资	国民生产总值	政治行动委员会公司比率
有效公司税率	1	100.0	0.0	0.0	0.0
	2	98.5	0.2	0.7	0.6
	3	93.2	0.4	0.7	5.8
	4	89.5	0.7	2.4	7.4
	6	83.5	5.0	2.7	8.8
	8	77.7	5.4	8.1	8.8
	10	74.2	5.7	11.7	8.5
	12	68.9	9.8	13.2	8.1
	16	63.4	14.7	13.2	8.8
国民生产总值	1	0.0	99.9	0.0	0.0
	2	5.4	90.9	1.5	2.3
	3	8.4	85.5	3.2	3.0
	4	10.9	83.6	2.8	2.7
	6	17.7	77.1	2.5	2.8
	8	25.2	69.8	2.5	2.6
	10	32.3	63.0	2.2	2.4
	12	36.7	59.0	2.1	2.3
	16	40.1	54.1	3.6	2.2
投资	1	0.0	39.6	60.3	0.0
	2	1.8	57.5	40.7	0.1
	3	4.6	66.1	28.4	0.9
	4	4.8	70.3	23.7	1.2
	6	6.7	73.4	17.6	2.4
	8	9.6	71.4	15.5	3.6
	10	13.9	68.0	14.6	3.5
	12	16.3	66.3	14.0	3.4
	16	17.0	65.5	14.0	3.4

续表

预测误差	创新				
	k	有效公司税率	投资	国民生产总值	政治行动委员会公司比率
政治行动委员会公司比率	1	4.0	2.5	4.4	89.1
	2	8.4	2.6	2.5	86.5
	3	12.0	5.0	1.2	81.8
	4	17.6	7.9	2.7	71.9
	6	20.4	12.0	5.3	62.3
	8	18.8	18.5	5.2	57.5
	10	17.6	24.6	4.9	53.0
	12	19.4	25.4	4.8	50.4
	16	22.8	26.8	5.0	45.3

注：表中报告的是创新导致的预测误差的百分比。数字表示由于列变量的创新导致的行变量在 k 季度之前的预测误差的百分比。

在表 4.12 中有四部分的预测方差百分比，每一个部分针对系统中一个变量。在第一个部分中，对税率的预测误差被归于其他 4 个变量（请注意，每一行对应的预测范围的和是 100%），所以在外生冲击发生的最初阶段，税率的所有预测误差都归于自身，这与冲击反应的考利斯基分解中变量的顺序一致。随后，我们从第 2 个季度跟踪至第 16 个季度，来看税率预测方差究竟有多少来自其自身，又有多少来自其他变量。在第 12 个季度中，我们看到，税率的总方差中有 30%来自其他 3 个变量，其中投资占 9.8%，国民生产总值占 13.2%，商业性政治行动委员会比率占 8.1%。

对于国民生产总值和投资来说，投资变量的创新解释了绝大部分的预测方差。从第 12 个季度到第 16 个季度，国民生产总值超过一半的方差是投资所导致的，而 40%是税率的创新所导致的。因此，国民生产总值主要是对这两个变量进行反应。对于投资的预测误差，我们看到，大部分是被其自身解释，投资变量中超过 65%的方差是被自身过去 12 个季度的创新所解释。

从上述重新计算的结果中，我们可以得出一些关键性的结论。首先，税率的冲击导致了投资模式的变化，因为税率的创新解释了过去 16 个季度的投资中

17%的方差。其次，税率的创新对委员会比率的影响也很大，它解释了 23%的方差。我们在之前已经运用外生性检验证明了税率并不是其他变量导致的，是外生于系统的。所以甚至在 16 个季度之后，税率的预测误差中超过 2/3 是由于自身的创新导致的。上述结论支持了命题 1、命题 2 和命题 3。

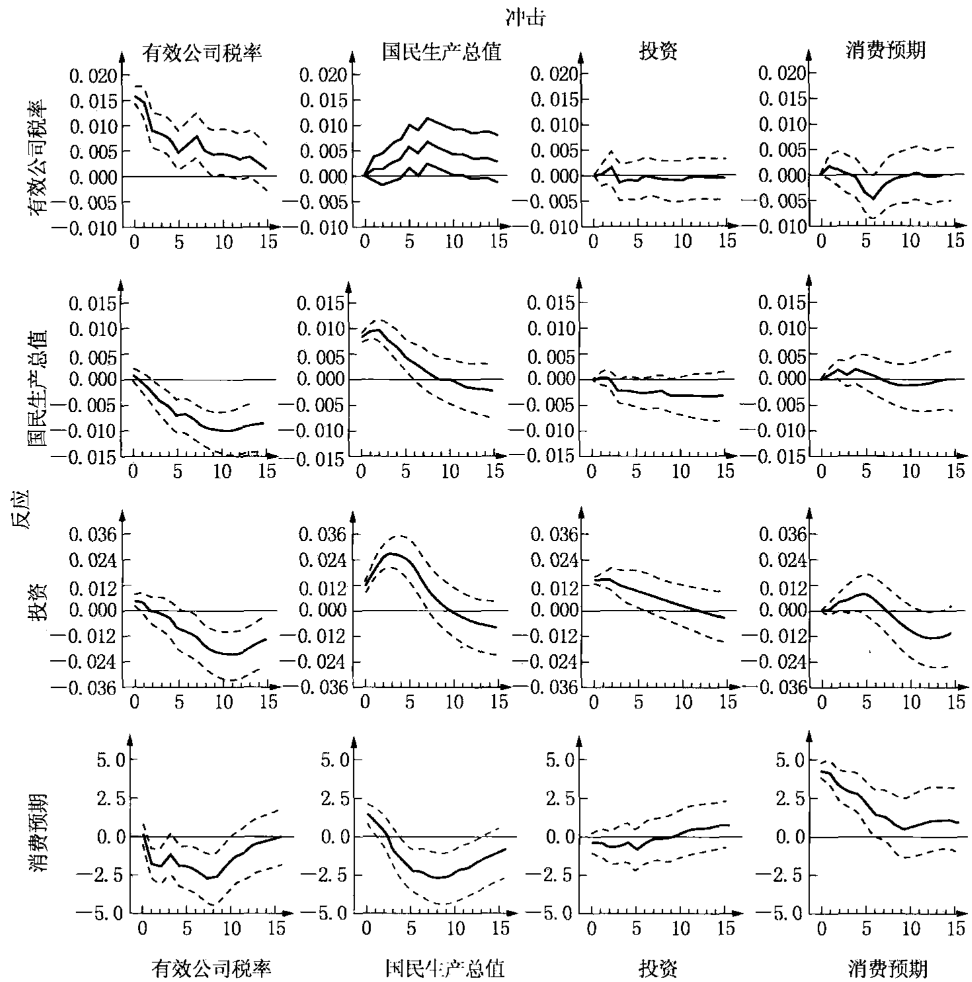
进一步的稳健性检查

威廉姆斯和柯林斯的一个贡献在于，他们解释了税率能够很好地预测未来的经济事件。本文的经验研究也很好地证明了这一点：税率的变化可以预测公司的投资和国民生产总值。如果税率和最终的税收政策能预测投资和经济增长，那么我们认为它也能预测消费预期。威廉姆斯和柯林斯指出，公众对经济走向的预期反应为消费预期，可以被其他变量运用向量自回归模型进行预测。

对上述分析进行稳健性检验的一个原因是，委员会比率这一变量的测量是从 1976 年以后才开始的，但是消费预期的指数是由密歇根大学从 1953 年就开始收集的。威廉姆斯和柯林斯是运用一个新的包含 4 个变量的向量自回归模型来评估税率对投资和国民生产总值的预测效度。该模型运用了税率、国民生产总值的对数、利率以及消费预期的对数。该分析加入了政治行动委员会这一变量，所以我们无法分析命题 1。但是由于这 4 个变量的数据可供使用的时间范围较大，所以可以用来评估税率和消费预期之间的关系。该模型也被用来验证之前模型中命题 2 和命题 3 的稳定性。

我们不再列出整个格兰杰因果分析的结果(Williams & Collins, 1997：表 5)，在这里只呈现移动平均反应过程。图 4.6 是含有 6 个滞后项的 4 个新变量的模型的移动平均过程反应图。

左上方的 3×3 个图与图 4.5 中相应位置的图是一样的。该图是运用以下几种方法来检验之前模型的稳健性：第一，运用更长的时间范围来进一步说明税率和投资之间的关系。图形说明，税率的升高会降低投资，与命题 2 相符。对税率冲击的 1.5%所引发的长期效应会导致投资降低 1%。其次，我们可以看到，税率是影响消费预期的，税率中每 1.5%的冲击会导致 6 个季度后的消费预



注：移动平均反应的时间范围是16个季度。

图4.6 4个变量在90%误差范围的移动平均分析(1953—1994年)

期发生2.5的降低，且统计上显著。相反，对消费预期的冲击却没有激起税率的反应，因为其置信区间包含0。这些结果与威廉姆斯和柯林斯报告的格兰杰因果检验结果一致。

向量自回归分析的结果支持威廉姆斯和柯林斯的命题。估计结果在不同样本、滞后时长和变量测量方式下都十分稳健。最后，稳健性检查的部分结果也支持税率会影响总体经济情况和消费预期的论断。

结论

我们用两个例子说明了基本的向量自回归分析。这些方法可以被用于具有更多变量的方程系统中来分析更为复杂的动态关系。我们的主要目的是，通过例证来告诉大家如何运用向量自回归模型来回答理论中提出的动态关系。

另外，必须说明，我们在这里所呈现的内容只是多元时间序列建模方法的一小部分。向量自回归模型、结构方程模型和误差纠正模型都被广泛地运用于各种动态建模、预测和政策分析。

还有一些对现有方法的拓展，例如我们提到过的结构向量自回归模型。在这个模型中，我们先在理论的指导下设定模型中的同期创新，这点与结构方程模型极为相似。换句话说，就是用结构方程模型中的方法来建构结构向量自回归模型。该模型可以被用于分析宏观供需关系(Blanchard & Quan，1989)、宏观经济状况(Bernanke，1986；Sims，1981、1986a)、货币政策(Leeper et al.，1996)以及描述总统支持率和宏观经济的关系(Williams，1990)。瓦格纳和扎正在进行模型的理论推进(Waggoner & Zha，2003)。

最后，尽管我们运用一个向量自回归模型来分析多元时间序列数据，但其目标在于不同时间序列之间的动态性和内生关系。本书已经示范了一些描述时间序列之间动态关系的主要方法，还提及了一些相关的模型，例如同期方程模型、误差纠正模型和结构方程模型。无论运用哪种方法，目的都是一样的，即回答不同时间序列之间的动态关系是什么样的？变量之间是不是彼此内生的？

附录 | 多元时间序列分析的软件

标准的回归软件也可以用于分析多元时间序列数据。如果想对自回归整合移动平均模型或者同时方程模型进行估计,许多统计软件都包含相关程序和假设检验的功能。但是,大多数从事时间序列分析的研究者很快会发现,这些常用的回归软件对识别时间序列数据的性质来说效果很差。因此,我们有必要在这里列出几条选择分析软件的意见。

进行一项时间序列数据分析时,需要知道的第一件事是,我们想要用时间序列数据做不同的事情(Brandt, 2002)。

第一,我们想要通过时间序列的子样本的组合来构建符合我们理论问题的时间序列数据。我们经常遇到需要把收集的数据切分成不同时间段的情况。尽管这一工作可以运用相关的统计软件来做,但是对数据和日期的设定十分关键,这需要软件能够识别带有时间序列性质的数据中的变量,并且知道如何将数据分组。好的时间序列软件能够较好地完成这一工作。

第二,时间序列软件能够识别、建构和控制具有时间单位的数据。这就要求软件中含有产生变量提前值或者滞后值的函数。另一个函数方面的重要要求是,它必须能够将二位计数的年份转换成四位计数的年份(如果数据同时包括 20 世纪和 21 世纪的年份)。有关时间和日期单位的工作可以由具有时间序列绘图功能的软件完成。

第三,对于结构方程和向量自回归模型来说,假设检验以及模型的拟合是与标准回归模型不同的。好的时间序列软件能出色地完成对自回归、格兰杰因果、滞后时长选择、单位根、误差纠正模型等问题的检验和诊断。

第四,只有时间序列软件能够胜任对一些复杂模型的设定和计算。我们也

可以运用多元线性回归的软件来对向量自回归模型和多阶段同时方程模型进行估计。但是模型中的动态模拟、冲击乘子、冲击反应效应、创新计算、蒙特卡洛模拟和误差范围的建立等问题只能通过专门的时间序列软件来完成。因此，我们应该花时间来甄别什么软件和程序语言是进行时间序列分析所需要的。

最后，根据我们以往教学、分析和发展多元时间序列模型的经验，我们认为，一个好的时间序列软件能够提供更快捷、更有效和更少错误的分析。

我们的主要目的并不是建议"某一个"程序或者软件包，而是指出被用于分析的软件应该具有以下一些功能。如果所用的软件在分析某些问题时存在缺陷，那么就不能用该软件进行此类分析，而是换用其他更合适的软件来完成这些工作。如果想对向量自回归模型的冲击反应进行假设检验，那么我们就必须选择能够对冲击反应、创新计算以及这些问题的各种例外情况进行分析的软件。

从我们的经验来看，有很多统计分析软件可被用于时间序列模型的分析。以下是几条挑选软件的基本标准：第一，所用的软件必须能够识别多元时间序列模型，并且包含分析向量自回归模型和结构方程模型的基本工具。第二，软件中含有一些专门的时间序列分析的模块。这些程序包括专门的假设检验和模型拟合方法。最后，软件必须包含较高水平的程序语言，从而能够用于专门的时间序列模型。因为时间序列模型中包含许多十分复杂的代数矩阵和时间序列函数，所以在这里，我们对一些软件进行一个简要的概览。

第一组软件包括 SPSS、SAS 和 Stata。这些基本的程序被广泛地用于社会科学的研究中。它们都可以完成一些时间序列分析。SAS 和 Stata 包括了很多专门估计向量自回归模型、冲击反应、创新计算、滞后时长以及格兰杰因果检验的程序。在 Stata 中，"help var"命令可以提供模型解释的一些说明。在 SAS 中，VARMAX 程序可以估计向量自回归模型、贝叶斯向量自回归模型以及向量误差纠正模型。SPSS 在这方面相对较弱，只能分析一些简单的同时方程模型。

还有一些软件程序是专门为时间序列分析而设计的。它们含有图形界面，需要编辑命令或允许不同模式之间的互动。Estima's RATS(时间序列回归分析)程序和 CATS(时间序列协同整合分析)程序能够为多元时间序列模型提供

广泛的分析。在向量自回归模型分析软件中，运用最广泛的那些程序都是汤马斯·多恩(Thomas Doan)写的。RATS 软件十分强大，但是该软件没有图形界面，所有的分析都必须通过自己编写命令来完成。该软件可以对向量自回归模型中的所有情况进行估计，也可以通过简单的矩阵程序软件来分析自己设定的模型。

JMulti 是另一个可用于分析向量自回归和向量误差纠正模型的软件。卢克波尔和克拉茨的书对该软件有详细的描述(Lutkepohl & Kratzig, 2004)。它可以通过冲击反应和创新计算对模型提供广泛的假设检验和解释。Eviews 是一个同时包含命令界面和图形界面的时间序列分析软件，可被用于分析单变量和多变量时间序列分析。最后，Ox 和 Oxmetric 软件也可以被用于上述分析。Ox 既可以通过图形界面进行分析，也允许分析者自己按照程序语言来编辑命令，从而对时间序列模型提供广泛的分析。

最后一组软件是一些具有高水平的程序语言和统计分析的软件，包括对代数矩阵和统计分析的 Aptech's Gauss 程序，还有 Insightful's Splus 软件，我们还可以用图形界面或者强大的 S 语言来进行统计和绘图分析。这些强大的程序包括“图书馆”、“软件包”和“工具栏”，能够为向量自回归和向量误差纠正模型提供很好的分析。这些高水平的程序语言需要分析者自己编写命令。然而，这些强大的功能也要付出一些代价，即它没有提供详细的时间序列函数，这就意味着，很多程序都需要分析者自己进行编写。

有关软件的其他信息可以很轻易地在互联网上找到。所有上述软件都可以在微软 Windows、苹果 OSX 和 Unix/Linux 操作系统中运行。

注释

[1] 一个显著的例外是经济学中的供求关系模型。

[2] 基本的要求是，方程中外生变量或者先决变量的数量一定要大于内生变量减去 1。

[3] 这与结构方程模型识别的标准定义一致(Judge, Griffiths, Hill, Lutkepohl & Lee, 1985: 573—574)。

[4] 弗里曼对该观点进行了具有说服力的讨论(Freedman, 1989)。

[5] 这里的重点是，与线性回归或同时方程模型的使用者类似，矩阵右边的变量必须是“满秩”的，或者符合对阶和秩的限定条件。尽管在实际操作中，模型即使满足了这些秩和阶的要求，也可能出现无法识别的情况，或者可能与其他的表述形式等同，即参数并不是变量的唯一表达形式。在这种情况下，我们就需要借助一些准则来完成模型的识别。

[6] 一般来说，时间序列整合到 d 阶可以被记作 $I(d)$，表示要进行 d 次求差来保证数据的稳定性。同样的 d 阶时间序列可以一起被放入模型。

[7] 我们也可以纳入一个移动平均成分或者建构一个向量自回归移动平均模型(详见 Lutkepohl, 2005)。

[8] $(m\times n)$ 矩阵 A 和 $(p\times q)$ 矩阵 B 的克罗尼克乘积是下列 $(mp)\times(nq)$ 矩阵：

$$A\otimes B=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

[9] 这些推导的最大似然方法见汉米尔顿的著作(Hamilton, 1994)。

[10] 这点很重要：去除季节性的数据可以用自回归移动平均方法或者 $X11$ 过程来放大季节性的序列相关。在向量自回归模型中加入足够多的滞后项，可以帮助我们正确地把握这些序列相关的模式。

[11] 运用过多彼此高度相关的滞后项可能会导致 $(X'X)$ 矩阵不可逆，也就是说，无法计算最小二乘估计。

[12] 请注意，对似然比检验的计算和似然函数不同，我们不必对误差协方差矩阵 Σ 进行转置。这是因为下列简化形式(运用行列式和对数形式的性质)：

$$\begin{aligned}2(L(\hat{\Sigma}, B, p_1)-L(\hat{\Sigma}, B, p_0)) &= 2\left[\frac{T}{2}\log|\hat{\Sigma}_1^{-1}|\frac{T}{2}\log|\hat{\Sigma}_0^{-1}|\right] \\ &= T\left(\log\left(\frac{1}{\hat{\Sigma}_1}\right)-\log\left(\frac{1}{\hat{\Sigma}_0}\right)\right) \\ &= -T(-\log(\hat{\Sigma}_1)+\log(\hat{\Sigma}_0)) \\ &= T(\log(\hat{\Sigma}_0)-\log(\hat{\Sigma}_1))\end{aligned}$$

[13] 对乘数检验进行调整的典型做法是伯努利 P 值调整或者瑟达克调整 P 值。对序列中第 k 个滞后值进行检验的伯努利调整值是 $p_i^b=\min(1, kp_i)$，其中 k 是假设检验中滞后项的个数，p_i 是未调整的 P 值(假设为 0.05)。相应的瑟达克调整是 $p_i^s=1-(1-p_i)^k$。

[14] 关于这一点，我们可以在 $\log|\hat{\Sigma}|$ 不会发生变化的假设下通过对比 AIC($p+1$) 和 AIC(p)来实现。两者的差别将是 m^2 的倍数。所以，$\log|\hat{\Sigma}|$ 从 $p+1$ 到 p 个滞后值的变化必须比更多的参数带来的惩罚因子大 m^2，检验结果才会支持更为简化的模型。

[15] 平方矩阵的迹是其对角线所有元素的加总。

[16] 另一种方法请见西姆斯的著作(Sims, 1972)以及弗里曼的著作(Freedman, 1983)。

[17] 本节是从布兰特和弗里曼讨论评估向量自回归模型动态性的不确定性的文章中节选出来的。

[18] 对蒙特卡洛抽样方法的介绍,请参见穆尼的著作(Mooney, 1997)。

[19] 误差纠正模型很少用于稳定时间序列数据,因为如果没有随机趋势,出现虚假回归的风险也很低。但是,许多经济和社会变量存在单位根和随机趋势的性质,所以在模型中有这类变量时,应该将其当做单位根过程处理。

[20] 这一表述方式与多于两个变量的向量误差纠正模型一样,因为 y_t 可以用于多于两个变量的矩阵。

[21] 参见 DeBoef & Granato, 1997。

[22] 例子中使用的数据和 RATS 软件命令可以在作者的网站上找到。

[23] 埃里克森等人提供了一系列动态模拟和进行误差纠正的方法来分析这些变量(Erikson, 2002)。

[24] 另外,如果我们认为宏观政党参与这个非负取值的变量在理论上不可能存在单位根,那么就可以省略对该变量的检验。

[25] 如果严格遵循赤池信息准则来对模型进行选择,那么就会导致最终选择的模型包括太多的滞后项(Lutkepohl, 2004)。

[26] 在后文中,我们加入两个滞后项,其结果是一样的。

[27] 对于较长的时间范围,我们可以只报告主要时期的预测误差方差的分解结果。

[28] 改变变量的顺序会改变考利斯基分解的标准化以及在计算移动平均反应时方程的顺序。

[29] 运用同样的数据对单位根和可能的协同整合关系进行讨论的内容,参见 Inclan, Quinn & Shapiro, 2001。对于运用这一数据所做的其他向量自回归分析,请见弗里曼等人的著作(Freeman et al., 1998)。

[30] 运用伯努利和瑟迪克 P 值调整方法时,我们还运用了 8 个滞后项的模型。但是结果和我们先前所列出的是一样的,因此我们选择较为简洁的模型。

[31] 双变量向量自回归模型产生的假设检验结果具有渐进正确性。如果出现了单位根,那么我们可以对求一阶差以后的数据运用向量自回归模型进行评估。

[32] 请注意,这一结果与威廉姆斯和柯林斯的结果略有不同,他们错误地将 8 个滞后项的方程估计结果报告为 4 个含有滞后项的方程估计结果。

[33] 这些结果和威廉姆斯和柯林斯的结果类似,但是并不完全相同。他们对于预测误差方差的分解是基于时变贝叶斯向量自回归模型的,但是本书的结果是基于非限定性的,不优先考虑贝叶斯过程的非时变的向量自回归结果。

参考文献

Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. F. (1993). *Co-integration, error correction and the econometric analysis of non-stationary data*. Oxford, UK: Oxford University Press.

Bernanke, B. (1986). "Alternative explanations of the money-income correlation." In *Carnegie-Rochester conference series on public policy*. Amsterdam: North-Holland.

Blanchard, O., & Quah, D. (1989). "The dynamic effects of aggregate demand and supply disturbances." *American Economic Review*, *79*, 655—673.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco: Holden Day.

Box-Steffensmeier, J. M., & Smith, R. M. (1996). "The dynamics of aggregate partisanship." *American Political Science Review*, *90*(3), 567—580.

Brandt, P. T. (2002). "Using the right tools for time series data analysis." *Political Methodologist*, *10*(2), 22—26.

Brandt, P. T., & Freeman, J. R. (2006). "Advances in Bayesian time series modeling and thestudy of politics: Theory testing, forecasting, and policy analysis." *Political Analysis*, *14* (1), 1—36.

Clarke, H. D., Ho, K., & Stewart, M. C. (2000). "Major's lesser (not minor) effects: Prime ministerial approval and governing party support in Britain since 1979." *Electoral Studies*, *19*, 255—273.

Cooley, T. F., & LeRoy, S. F. (1985). "Atheoretical macroeconomics: A critique." *Journal of Monetary Economics*, *16*, 283—308.

Cromwell, J. B., Labys, W. C., Hannan, M. J., & Terraza, M. (1994). *Multivariate tests for time series models*. Thousand Oaks, CA: Sage.

DeBoef, S., & Granato, J. (1997). "Near inte grated data and the analysis of political relationships." *American Journal of Political Science*, *41*(2), 619—640.

Doan, T., Litterman, R., & Sims, C. (1984). "Forecasting and conditional projection using realistic prior distributions." *Econometric Reviews*, *3*, 1—100.

Dolado, J. J., & Lutkepohl, H. (1996). "Making Wald tests work for cointegrated VAR systems." *Econometric Reviews*, *15*(4), 369—386.

Engle, R. F., & Granger, C. W. J. (1987). "Co-integration and error correction: Representation, estimation and testing." *Econometrica*, *55*, 251—276.

Erikson, R. S., MacKuen, M. B., & Stimson, J. A. (2002). *The macropolity*. New York: Cambridge University Press.

Freeman, J. R. (1983). "Granger causality and the time series analysis of political relationships." *American Journal of Political Science*, *27*(3), 327—358.

Freeman, J. R., Williams, J. T., Houser, D., & Kellstedt, P. (1998). "Long memoried processes, unit roots and causal inference in political science." *American Journal of Political Sci-*

ence, *42*(4), 1289—1327.

Freeman, J. R., Williams, J. T., & Lin, T.-M. (1989). "Vector autoregression and the study of politics." *American Journal of Political Science*, *33*, 842—877.

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica*, *37*, 424—438.

Granger, C. W. J., & Newbold, P. (1974). "Spurious regressions in econometrics." *Journal of Econometrics*, *2*, 111—120.

Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series* (2nd ed.). San Diego, CA: Academic Press.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Harvey, A. C. (1990). *The econometric analysis of time series* (3rd ed.). Cambridge, MA: MIT Press.

Hatanaka, M. (1975). "On the global identification of the dynamic simultaneous equations model with stationary disturbances." *International Economic Review*, *16*(3), 545—554.

Inclan, C., Quinn, D. P., & Shapiro, R. Y. (2001). "Origins and consequences of changes in US corporate taxation, 1981—1998." *American Journal of Political Science*, *45*(1), 179—201.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford, UK: Oxford University Press.

Judge, G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., & Lee, T.-C. (1985). *Theory and practice of econometrics* (2nd ed.). New York: Wiley.

Kilian, L. (1998). "Small-sample confidence intervals for impulse response functions." *Review of Economics and Statistics*, *80*, 186—201.

Krolzig, H.-M. (1997). *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*. Berlin: Springer.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root." *Journal of Econometrics*, *54*, 159—178.

Leeper, E. M., Sims, C. A., & Zha, T. (1996). "What does monetary policy do?" *Brookings Papers on Economic Activity*, *2*, 1—63.

Litterman, R. B., & Weiss, L. (1985). "Money, real interest rates, and output: A reinterpretation of postwar U. S. data." *Econometrica*, *53*(1), 129—156.

Lutkepohl, H. (1985). "Comparison of criteria for estimating the order of a vector." *Journal of Time Series Analysis*, *6*, 35—52. (Correction: 1987, 8, 373)

Lutkepohl, H. (1990). "Asymptotic distributions of impulse response functions and forecasterror variance decompositions in vector autoregressive models." *Review of Economics and Statistics*, *72*, 53—78.

Lutkepohl, H. (2004). "Vector autoregressive and vector error correction models." In H. Lutkepohl & M. Kratzig (Eds.), *Applied time series econometrics* (pp. 86—158). Cambridge, UK: Cambridge University Press.

Lutkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer.

Lutkepohl, H., & Kratzig, M. (Eds.). (2004). *Applied time series econometrics*. Cambridge,

UK: Cambridge University Press.

Lutkepohl, H., & Reimers, H.-E. (1992a). "Granger-causality in cointegrated VAR processes: The case of term structure." *Economics Letters*, *40*, 263—268.

Lutkepohl, H., & Reimers, H.-E. (1992b). "Impulse response analysis of cointegrated systems." *Journal of Economic Dynamics and Control*, *16*, 53—78.

Mills, T. (1991). *Time series techniques for economists*. Cambridge, UK: Cambridge University Press.

Mittnik, S., & Zadrozny, P. A. (1993). "Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic models." *Econometrica*, *20*, 832—854.

Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Ostrom, C. (1990). *Time series analysis: Regression techniques*. Thousand Oaks, CA: Sage.

Ostrom, C., & Smith, R. (1993). "Error correction, attitude persistence, and executive rewards and punishments: A behavioral theory of presidential approval." *Political Analysis*, *3*, 127—184.

Pagan, A. (1987). "Three econometric methodologies: A critical appraisal." *Journal of Economic Surveys*, *1*(1), 3—24.

Philips, C. B. (1986). "Understanding spurious regressions in econometrics." *Journal of Econometrics*, *33*, 311—340.

Reinsel, G. C. (1993). *Elements of multivariate time series analysis*. New York: Springer-Verlag.

Runkle, D. E. (1987). "Vector autoregressions and reality." *Journal of Business and Economic Statistics*, *5*, 437—442.

Saikkonen, P., & Lutkepohl, H. (1999). "Local power of likelihood ratio tests for the cointegrating rank of VAR processes." *Econometric Theory*, *15*, 50—78.

Saikkonen, P., & Lutkepohl, H. (2000a). "Testing for cointegrating rank of a VAR process with an intercept." *Econometric Theory*, *16*, 373—406.

Saikkonen, P., & Lutkepohl, H. (2000b). "Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process." *Journal of Time Series Analysis*, 21, 435—456.

Sims, C. A. (1972). "Money, income, and causality." *American Economic Review*, *62*, 540—552.

Sims, C. A. (1980). "Macroeconomics and reality." *Econometrica*, *48*(1), 1—48.

Sims, C. A. (1981). "An autoregressive index model for the U. S., 1948—1975." In J. Kmenta & J. B. Ramsey (Eds.), *Large-scale macro-econometric models* (pp. 283—327). Amsterdam: North-Holland.

Sims, C. A. (1986a). "Are forecasting models usable for policy analysis?" *Quarterly Review, Federal Reserve Bank of Minneapolis*, *10*, 2—16.

Sims, C. A. (19866). "Specification, estimation, and analysis of macroeconomic models." *Journal of Money, Credit and Banking*, *18*(1), 121—126.

Sims, C. A. (1987). "Comment [on Runkle]." *Journal of Business and Economic Statistics*, *5*(4), 443—449.

Sims, C. A. (1988). "Uncertainty across models." In *Papers and proceedings of the one-hundredth annual meeting of the American Economic Association* (pp. 163—167). Nashville, TN: Amer-

ican Economic Association American Economic Review.

Sims, C. A. (1996). "Macroeconomics and methodology." *Journal of Economic Perspectives*, *10*, 105—120.

Sims, C. A., Stock, J. H., & Watson, M. W. (1990). "Inference in linear time series models with some unit roots." *Econometrica*, *58*(1), 113—144.

Sims, C. A., & Uhlig, H. (1991). "Understanding unit rooters: A helicopter tour." *Econometrica*, *59*(6), 1591—1599.

Sims, C. A., & Zha, T. (1998). "Bayesian methods for dynamic multivariate models." *International Economic Review*, *39*(4), 949—968.

Sims, C. A., & Zha, T. (1999). "Error bands for impulse responses." *Econometrica*, *67*(5), 1113—1156.

Stimson, J. (1999). *Public opinion in America: Moods, cycles, and swings* (2nd ed.). Boulder, CO: Westview.

Toda, H. Y., & Phillips, P. C. B. (1993). "Vector autoregressions and causality." *Econometrica*, *61*(6), 1367—1393.

Toda, H. Y., & Yamamoto, T. (1995). "Statistical inference in vector autoregressions with possibly integrated processes." *Journal of Econometrics*, *66*, 225—250.

Waggoner, D. F., & Zha, T. (1999). "Conditional forecasts in dynamic multivariate models." *Review of Economic and Statistics*, *81*(4), 639—651.

Waggoner, D. F., & Zha, T. A. (2003). "A Gibbs sampler for structural vector autoregressions." *Journal of Economic Dynamics & Control*, *28*, 349—366.

Williams, J. T. (1990). "The political manipulation of macroeconomic policy." *American Political Science Review*, *84*(3), 767—795.

Williams, J. T. (1993). "What goes around comes around: Unit root tests and cointegration." *Political Analysis*, *4*, 229—235.

Williams, J. T., & Collins, B. K. (1997). "The political economy of corporate taxation." *American Journal of Political Science*, *41*(1), 208—244.

Williams, J. T., & McGinnis, M. D. (1988). "Sophisticated reaction in the U. S.-Soviet arms race: Evidence of rational expectations." *American Journal of Political Science*, *32*(4), 968—995.

Wold, H. (1954). *A study in the analysis of stationary time series* (2nd ed.). Uppsala, Sweden: Almqvist and Wiksell.

Zapata, H. O., & Rambaldi, A. N. (1997). "Monte Carlo evidence on cointegration and causation." *Oxford Bulletin of Economics and Statistics*, *59*(2), 285—298.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: Wiley Interscience.

Zellner, A., & Palm, F. C. (Eds.). (2004). *The structural econometric time series analysis approach*. Cambridge, UK: Cambridge University Press.

Zha, T. (1998). "A dynamic multivariate model for the use of formulating policy." *Economic Review (Federal Reserve Bank of Atlanta)*, *83*(1), 16—29.

译名对照表

Autoregressive Distributed Lag Models(ADL)	分布滞后自回归
Autoregressive Integrated Moving Average(ARIMA)	自回归整合移动平均法
Box-Jenkins univariate time series analysis	博克斯—詹金斯单时间序列分析
Breusch-Godfrey Lagrangian multiplier	布雷施—戈弗雷拉格朗日乘子法
Cholesky decomposition	考利斯基分解
cointegration	协同整合
Error Correction Models(ECMs)	误差修正模型
factor analysis	(动态)因子分析
impulse response analysis	冲击反应分析
Impulse Response Function(IRF)	冲击反应函数
innovative accounting	创新计算
Moving Average Response Analysis(MAR)	移动平均反应分析
multivariate seemingly unrelated regression model	多元相依回归
portmanteau test	混数检验
seemingly unrelated regressions	似乎不相关回归
simultaneous equation model	同时方程模型
Simultaneous or Structural Equation Systems(SEQ)	同时方程或结构方程系统
Vector Autoregression(VAR)	向量自回归
Wald decomposition theorem	沃德分解定理

Cohort Analysis(2nd edition)
Copyright © 2005 by SAGE Publications, Inc.
All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.
This simplified Chinese edition for the People's Republic of China is published by arrangement with SAGE Publications, Inc. © SAGE Publications, Inc. & TRUTH & WISDOM PRESS 2011.
本书版权归 SAGE Publications 所有。由 SAGE Publications 授权翻译出版。
上海市版权局著作权合同登记号:图字 09-2009-549

Longitudinal Research(2nd edition)
Copyright © 2002 by SAGE Publications, Inc.
All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.
This simplified Chinese edition for the People's Republic of China is published by arrangement with SAGE Publications, Inc. © SAGE Publications, Inc. & TRUTH & WISDOM PRESS 2011.
本书版权归 SAGE Publications 所有。由 SAGE Publications 授权翻译出版。
上海市版权局著作权合同登记号:图字 09-2009-552

Analyzing Repeated Surveys
Copyright © 1997 by SAGE Publications, Inc.
All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.
This simplified Chinese edition for the People's Republic of China is published by arrangement with SAGE Publications, Inc. © SAGE Publications, Inc. & TRUTH & WISDOM PRESS 2011.
本书版权归 SAGE Publications 所有。由 SAGE Publications 授权翻译出版。
上海市版权局著作权合同登记号:图字 09-2009-551

Multiple Time Series Models
Copyright © 2007 by SAGE Publications, Inc.
All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.
This simplified Chinese edition for the People's Republic of China is published by arrangement with SAGE Publications, Inc. © SAGE Publications, Inc. & TRUTH & WISDOM PRESS 2011.
本书版权归 SAGE Publications 所有。由 SAGE Publications 授权翻译出版。
上海市版权局著作权合同登记号:图字 09-2009-550